Implicit User Profiling for On Demand Relevance Feedback

Yoshinori Hijikata
Osaka University
1-3 Machikaneyama
Toyonaka OSAKA 560-8531, JAPAN
hijikata@sys.es.osaka-u.ac.jp

ABSTRACT

In the area of information retrieval and information filtering, relevance feedback is a popular technique which searches similar documents based on the documents browsed by the user. If the user wants to conduct relevance feedback on demand, which means the user wants to see similar documents while reading a document, the existing user profiling techniques cannot acquire keywords in high precision that the user is interested in at such a short time. This paper proposes a method for extracting text parts which the user might be interested in from the whole text of the Web page based on the user's mouse operation in the Web browser. The objective of this research is to (1) find what kind of mouse operation represent users' interests, (2) see the effectiveness of the found mouse operation in selecting keywords, and (3) compare our method with tf-idf, which is the most fundamental method used in many user profiling systems. From the user experiment, the precision to select keywords of our method is about 1.4 times compared with that of tf-idf.

Categories and Subject Descriptors: H.5.4 [Hypertext/ Hypermedia]: User issues
General Terms: Experimentation, Human Factors
Keywords: keyword selection, relevance feedback, mouse operation, tf-idf

1. INTRODUCTION

There are many search engine services on the Web that support users in acquiring their target information. When the user inputs some keywords as a search key, the search engine recommends pages that include the input keywords. The number of pages accessible by search engines has passed one billion pages[1]. Technologies for narrowing the number of search results are regarded as important, and many researchers have been working on these technologies. Methodology that relieves users from studying special knowledge about search engines on the Web is important, because there are many kinds of users on the Web. Relevance feedback[2] is one such method.

Relevance feedback (i) asks the user to indicate pages most relevant to his/her interests from the search results, and (ii) searches again using keywords specific to those selected pages. Generally, the selection of keywords is done from the pages returned as search results by the search engines, and this method selects new keywords from the complete text of those pages. Therefore this method has a problem in that not all the selected keywords have to do with the user's interests[5]. Another problem is that it takes a lot of effort by the users to indicate suitable pages.

User Profiling, which gathers the information about the user's interest, is important for this problem. Many user profiling techniques have been studied in the field of information filtering system or recommender system [3, 4]. However the construction of the user profile in most techniques extends over a long period. This is because they construct the user profile by using many keywords collected from many pages. This means it is impossible to select keywords in high precision at the short time like the user is staying in a page. Therefore most user profiling techniques cannot be used in selecting keywords for on demand relevance feedback like this.

In this paper, as a solution for the first problem in relevance feedback, we propose using only the parts that the user might be interested in, instead of using the entire pages. As a solution for the second problem, we propose using the user's browsing operations to determine his/her interests instead of asking the user to explicitly indicate the pages (or the parts in a page) that he/she had an interest in. We focus on the situation that the user uses a mouse as an input device while browsing Web pages and solve the abovementioned problems by the following method (after here we call this method "mouse-based method"):

1. Extract operations that might occur because of the user's especial interest from the user's ordinary mouse operations while browsing pages.
2. Extract by sentence or line the text parts that are the targets of those extracted operations.

We expect the following results by using this method:

1. The system can automatically find the keywords relevant to the user's interests without requiring any special efforts by the user.
2. The system can eliminate many noise keywords, the keywords unrelated to the user's interests, from the texts used for relevance feedback.

IUI'04, January 13–16, 2004, Madeira, Funchal, Portugal.

Furthermore the most important advantage is that our method can construct the user profile at the very short time because it narrows down the text part in a page which the user apparently indicates with the mouse pointer. The objective of this research is to (1) find what kind of mouse operation represent users' interests, (2) see the effectiveness of the found mouse operation in selecting keywords, and (3) compare our method with tf-idf, which is the most fundamental method used in many user profiling systems.

2. RELATED WORK

There are two basic approaches for user profiling[5, 6].

1. Explicit (Direct) method:

This method acquires user profiles by (i) asking users to answer preliminary questionnaires about topics or keywords which they are interested in, or (ii) asking users to grade the pages they have browsed for interest and relevance. Ringo[7] and SIFT[8] use the former approach. GroupLens[9], Syskill & Weber[10], News-Weeder[11], ClixSmart[12] and AntWorld[13] use the latter approach. The advantage of this method is that it is reliable because it acquires the user profiles directly from the users. However these approaches also have some disadvantages. Generally, completing a preliminary questionnaire sufficiently detailed to allow a user to adequately describe his/her interests as keywords is a troublesome task, and grading pages also takes a lot of efforts from the users. Method (ii) also has a problem that it selects keywords from the whole text of the page and the selected keywords include many that the user is not interested in.

2. Implicit (Indirect) method:

This method acquires user profiles by estimating the users' degree of interest in the pages the users have browsed based on such factors as (i) the time spent reading the pages (browsing time)[14] or (ii) the specific mouse button operations or the scroll operations performed while reading the pages[15], or (iii) the user's eye mark while reading pages[16, 17]. The advantage of this method is that it does not require any mental efforts by the users. One of the problems with method (i) is that the system usually cannot know when the user opens a page and then starts doing some other work or leaves the PC because the browsing time is usually acquired in the server. Existing research on method (ii) monitors for such actions as when the user pushes a button for enlarging an article in a news system or when the user scrolls the window that displays the article. Detecting these operations allows the system to judge whether the user was interested in the entire page. However the system cannot always locate which part of the page the user was interested in from these operations. Method (iii) has a possibility to specify the text part that the user was interested in. However it leaves the problem of the special equipment or device to recognize the user's eye mark.

Many user profiling systems also use tf-idf weights [2] for keyword selection. The tf-idf approach weights keywords based on each keyword's appearance frequency in the document and its appearance frequency in other documents. However tf-idf weights keywords based on the statistics of the entire document even if the user was only interested in a part of the document. Therefore some of the weights on keywords do not reflect the user's interests.

Our research (the mouse-based method) can be classified with the implicit methods because it estimates the users' interests based on the mouse operations. It differs from the existing research approaches in that (i) it estimates the user's interest from the ordinary mouse operations, including even the ones the user performs unconsciously, (ii) it extracts the parts the user might be interested in not by the page but by the sentence or line (This means that it is possible to construct the user profile at the short time), and (iii) it does not need a special device (see Table 1).

3. PRELIMINARY SURVEY

Kantor[13] reports that he discovered that users tend to follow the mouse pointer by the eye while browsing Web pages. As one of the reasons of the above-mentioned behavior, he pointed out that the user has to click links that he/she is interested in by the mouse on the Web. However he does not show what kind of operations performed by users while browsing Web pages and whether or not such operations have to do with their interests.

We surveyed characteristic operations which may occur according to users' interests. In this survey, we conducted observations about users' operations while they are browsing Web pages. In these observations, the users freely browsed Web pages they liked and the observer watched their mouse operations. 31 users participated in this survey as subjects. This survey detected the following characteristic operations. (We eliminated operations to directly specify the targets of the users' interests such as inputting some keywords that the user is interested in into the text field of a search engine.)

- Text tracing: Moving the mouse pointer along a sentence while reading.
- Link pointing: Positioning the mouse pointer on a link, but not clicking the link.
- Link clicking: Clicking on a link to move to another page.
- Text selection: Selecting text by dragging the mouse pointer.
- Scrolling: Scrolling a window at a certain speed.
- Bookmark registration: Registering a page as a bookmark.
- Saving: Saving an HTML document.
- Printing: Printing a page.
- Window movement: Moving a window of the Web browser.
- Window resizing: Changing the window size of the Web browser.

Some of the operations are necessary for browsing Web pages or using the Web browser's functions. The other operations are not necessary for browsing Web pages or using the Web browser's functions, but users perform them unconsciously. Out of these operations, the operations whose targets can be text are text tracing, link pointing, link clicking and text selection.

To judge whether or not we can use these four kinks of operations for extracting text parts, it is necessary to see how many users perform them. We observed the 20 users' operations during 10-minute browsing and counted the number of times that each operation occurred. Figure 1 shows the result. Although there is a variety in the number of times to perform according to the type of operation, we found that in every type of operation there are users who perform it. Therefore in this research, we will investigate whether or not the target text part of these four kinds of operations are actually the part the user was interested in by the experiment.

Table 1: Related works.

Method	Required for user's efforts?	Unit of interest	Required time	Required for special devices?
Preliminary questionnaire [7, 8]	Yes	—	Short	No
Page rating [9]-[13]	Yes	Page	Long	No
Browsing time [14]	No	Page	Long	No
Special button & scroll operation [15]	No	Page	Long	No
Eye mark [16, 17]	No	Text part	Short	Yes
Mouse-based method	No	Text part	Short	No

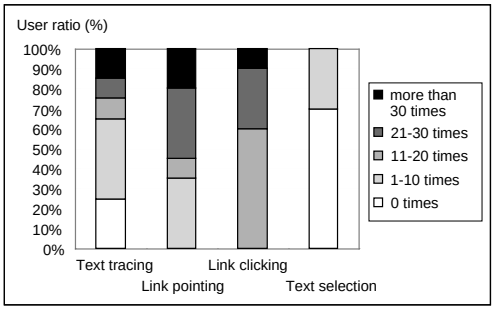

Figure 1: User ratio according to the number of operations.

4. EXPERIMENTAL METHOD AND SYSTEM IMPLEMENTATION

4.1 Experimental Method

Generally the unit of process in information retrieval and information filtering is keyword. In this experiment, we build a system which extracts the target text parts of the four types of operation found in the former section. We see whether or not keywords in the extracted text are actually the ones the user is interested in. The experimental method we used is as follows:

1. The subject searches for the Web sites he/she wants to browse in advance of the experiment.
2. When the experimental observations begin, the subject freely browses the selected Web sites.
3. Every time the subject moves from a page to another page, he/she answers a questionnaire about the previous page. In this questionnaire, the system displays all keywords extracted from the page, and the subject checks only keywords he/she was interested in (see Figure 2).
4. The experimenter compares the keywords checked by the subject and the keywords extracted according to the mouse operation. The experimenter determines the effectiveness by calculating some parameters.

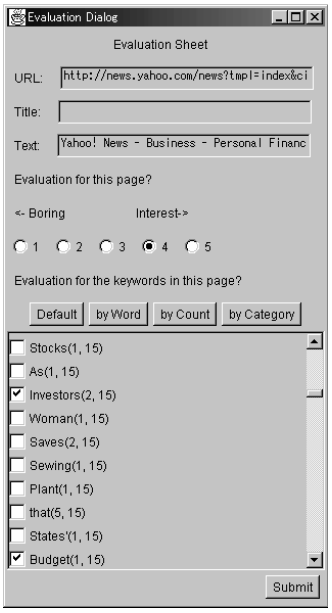

Figure 2: Questionnaire window.

4.2 System Implementation

We developed a system for the experiment which extracts text parts according to the mouse operation. The system is built in JavaScript and Java so that the user can use the Web browser that he/she usually uses. Figure 3 depicts the system's structure. For embedding the JavaScript and Java applet programs of the system into Web pages, we developed an embed proxy server. The embed proxy server also calls a morphological analyzer[18] and generates windows for questionnaires (see Figure 2) using output keywords. The extracted text parts are sent to the server and stored there.

The JavaScript program detects the user's operation event on the Web browser via DOM(Document Object Model)[19] interface. After that, it informs the Java applet program the event with other parameters such as coordinates of the mouse pointer in a fixed format. Examples of the mouse events with this format are as follows:

```
936332393593,blur,frames(0),7,BODY
936332407468,focus,frames(0),7,BODY
936332410218,mouseover,frames(0),7,BODY,215,0
936332410265,mousemove,frames(0),7,BODY,215,0
```

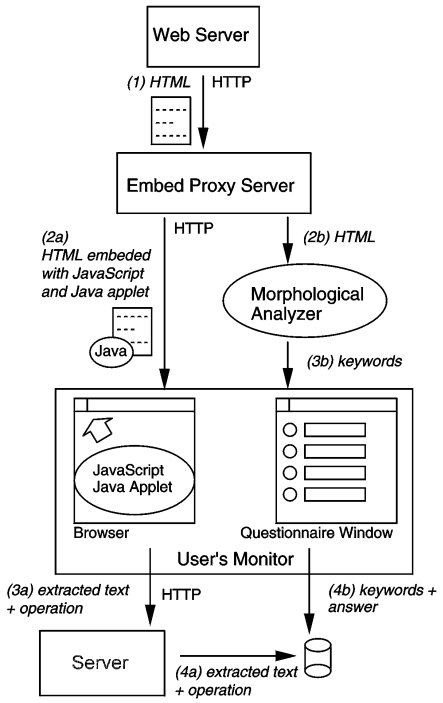

Figure 3: System structure.

4.3 Operation extraction and text extraction

The Java applet program extracts the four kinds of operations by analyzing the operation events and extracts the target text part of the operation. The actual method is as follows:

(1) Text tracing

First, the system detects continuous movement of the mouse pointer in a horizontal direction. For this detection, every time an onmousemove event occurs, the system calculates the angle of the mouse movement relative to the horizontal and the time between the current onmousemove event and the previous onmousemove event. (For calculating the angle, the system uses the current onmousemove event, and an onmousemove event that occurred n-times before.) If the angle is below a threshold Ar and the time is below a threshold Tr, the system regards the movement as a continuous movement in a horizontal direction.

Second, when the system detects such a movement of the mouse pointer, the system calculates the distance and velocity of the movement. If the distance is longer than a threshold L and the velocity is slower than a threshold V, the system regards this operation as a text tracing operation.

(2) Link pointing

The system regards the operation as a link pointing operation when an onmouseover event occurs on a link object, but there is no onclick event afterwards, and an onmouseout event occurs after a time Tp.

(3) Link clicking

The system regards an onclick event on a link object as a link clicking operation.

Table 2: Parameters for detecting operations.

Parameter	Value
Angle $A_r(\mid \tan\theta \mid)$	0.25
Time T_r	750(msec)
Length L	40(pixels)
Velocity V	0.45(pixels/msec)
History n	2
Time T_p	750(msec)

(4) Text selection

The system regards the operation as a text selection when a onmouseup event at the end of a onselect event occurs.

We use some parameters for extracting operations. We set the parameters heuristically by recording five users' operations while they are browsing Web pages and analyzing them (see Table 2). In actual text tracing operation found in the preliminary experiment and in the operation recording in this section, users do not strictly trace the line they are reading but just unconsciously move the mouse pointer to the right and in short distance. Therefore we set parameters to recognize the short mouse movement to the right as text tracing.

We use DynamicHTML[20] for extracting the text. When the system extracts the text which is the target of a text tracing operation, it also extracts the text which exists on the line above the line where the mouse pointer is. This is because of what we discovered from the observations. We found two cases in users' text tracing operations: (i) they move the mouse pointer on a straight line within the line where they are reading, or (ii) they move the mouse pointer on a line below the text line where they are reading.

4.4 User's Browsing in the Experiment

Five users (three women and two men in their twenties or thirties) participated in the experiment as subjects. We used data from 120 Web pages for the analysis. Table 3 shows the objective of each user's browsing, the characteristics of the pages that each user browsed, the average number of keywords in those pages, the average number of keywords the user checked as interesting ones in each page and the number of pages the user has browsed.

5. EVALUATION

5.1 Objective of Evaluation

This section sees whether or not the target text part of each type of operation is actually the part the user was interested in. Namely we will see whether or not the ratio of keywords that the user was interested in is higher in the target text part of each type of operation than in the whole text of the page. After that, we will compare our mouse-based method, which extracts keywords based on the four kinds of operations, with other keyword extraction methods. In this comparison, we will see the validity in text extraction of the mouse-based method by comparing with the method for extracting keywords at random, we call this "random extraction." We also compare the mouse-based method with tf-idf which is the most popular keyword selection method in information retrieval and information filtering.

Table 3: Browsing data for the experience.

User	The objective of browsing	The characteristics of the pages	NK	NC	NP
User A	Clicking a link of the mail magazine published by a news site on the mail software and browsing each news article.	Pages consisting of text and some figures. Figures are banner ads and photos for the article.	198	3.8	20
User B	Browsing a Web site for cars from its top page. Clicking a link of the mail magazine published by a news site on the mail software and browsing each news article.	A top page with many links and pages consisting of text and some figures. Figures are banner ads.	351	3.6	20
User C	Browsing personal sites from their top pages for essays and restaurant information.	Top pages with many links and pages consisting of text and some figures for the articles.	241	4.7	29
User D	Selecting Web sites for a popular singer in a commercial index service site and browsing concert information and bulletin boards in each site. Clicking a link of the mail magazine published by a news site on the mail software and browsing each news article.	A page with many links in an index service site, each site's top page with some links and figures, pages offering data as lists or tables, and pages of bulletin board. Few figures except for each site's top page.	157	1.1	25
User E	Browsing some personal or cities' Web sites offering travel information.	Top pages with some links and pages consisting of text and some figures. Some of the figures are large maps.	124	2.7	26

NK: Average number of keywords in a page.
NC: Average number of keywords checked by the user as interesting ones in a page.
NP: Number of pages the user browsed.

In this evaluation we will calculate the following three parameters: (1) keyword precision, (2) keyword recall, and (3) noise recall. Keyword precision is the ratio of the keywords that the user is interested in in relation to the extracted keywords. Keyword recall is the ratio of the extracted keywords in relation to the keywords that the user is interested in. Noise recall is the ratio of the extracted keywords in relation to the keywords that the user is not interested in (noise keywords). When we subtract noise recall from 1, we get the ratio of the reduced noise keywords in relation to the noise keywords in the page (noise reduction rate). Considering the usage in information retrieval and information filtering, keyword precision can evaluate the effectiveness of the extracted keywords. Using keyword recall and noise recall besides keyword precision, we can evaluate the effectiveness of the keyword extraction method. The equations to calculate these parameters are as follows:

- Keyword precision $=\mid B \mid / \mid A \mid$
- Keyword recall $=\mid D \mid / \mid C \mid$
- Noise recall $=\mid F \mid / \mid E \mid$

A, B, C, D, E and F in the above equations have the following meanings:

- A: The set of keywords extracted by the system.
- B: The set of keywords which are included in the set A and checked by the user as interesting ones.
- C: The set of keywords checked by the user as interesting ones in the whole text of the page.
- D: The set of keywords which are included in the set C and extracted by the system.
- E: The set of noise keywords in the whole text of the page.
- F: The set of noise keywords which are included in the set E and are extracted by the system.

5.2 Validity for Type of Operation

Figure 4 shows the keyword precision in every type of operation and the keyword precision in the whole text of the page. The keyword precision is higher in the extracted text than in the whole text for every user and for every type of operation. Figure 5 shows the number of times the user performed each type operation in a page. We can see there is individual difference in the frequency to perform the operation in text tracing and link pointing operation. The frequency to perform link click of User A is lower than other users. This is because User A read each new article by clicking the links of a mail magazine published by a news site on his mail software and hardly clicked links on the Web page. Although there was individual difference depending on the type of operation, we saw the extracted text part in every type of operation includes the keywords that the user was interested in at higher ratio than in the whole text of the page.

5.3 Comparison with Other Methods

Random extraction and tf-idf can extract keywords at any ratio. However the mouse-based method cannot extract keywords at a fixed ratio. Therefore we will calculate keyword narrowing rate which represents how much the mouse-based method narrows the text part from the whole text of the page. We will extract keywords at keyword narrowing rate of the mouse-based method also in random extraction and tf-idf. This means we will compare these methods when they extract keywords at the same ratio. The equation to calculate keyword narrowing rate is as follows:

Keyword narrowing rate $=\mid H \mid / \mid G \mid$

H and G in the above equation have the following meanings:

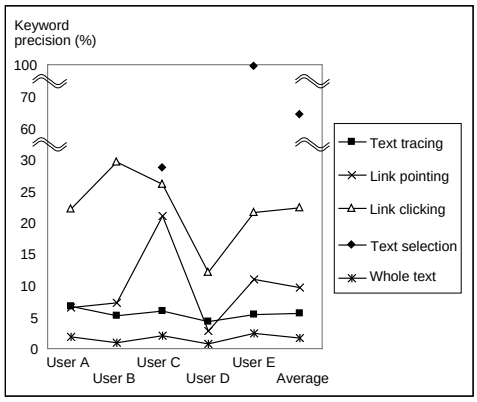

Figure 4: Keyword precesion for each type of operation.

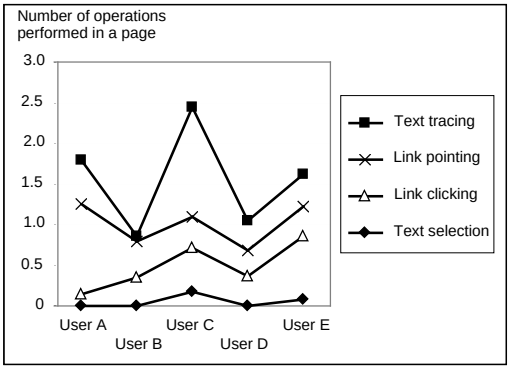

Figure 5: Number of performed operations per a page.

- G: The set of keywords in the whole text of the page.
- H: The set of keywords which are included in the set G and extracted by the mouse-based method.

tf-idf needs the document set defined in advance. In this experiment, we created vector spaces using keywords in all pages browsed by each user. We selected keywords based on the weights of tf-idf at the keyword narrowing rate.

Table 4 shows the keyword narrowing rate. Figure 6, 7, 8 shows the keyword precision, the keyword recall and the noise recall. Compared to random extraction, the keyword precision and the keyword recall of the mouse-based method are approximately four times on the average of all users. The difference of the noise recall between the mouse-based method and random extraction is small although the noise recall of the mouse-based method is slightly better than random extraction. This is because more than 98% of keywords in the whole text of the page are noise keywords (We can see this from that the average keyword precision of all users in the whole text of the page is less than 2%). Therefore the mouse-based method extracts more keywords the user was interested in than random extraction and reduces noise keywords at almost the same ratio of random extraction.

Compared to tf-idf, the keyword precision and the keyword recall of the mouse-based method is about 1.4 times

Table 4: Keyword narrowing rate.

User	Keyword narrowing rate(%)
User A	9.74
User B	3.50
User C	8.62
User D	7.76
User E	14.32
Average	8.78

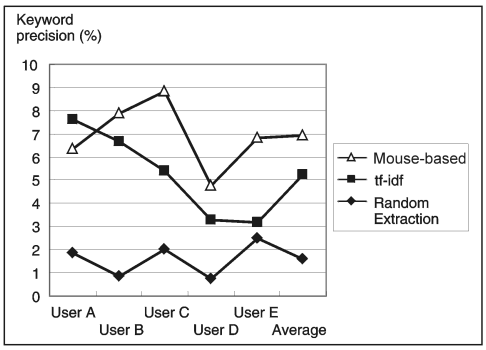

Figure 6: Keyword precision.

on the average of all users. For the users except for User A, the keyword precision and the keyword recall of the mouse-based method is better than those of tf-idf. There is difference in browsing behaviors between User A and User B-E. User A browsed only the pages that display one news article about IT (Information Technology). User B-E browsed various kinds of pages such as top pages of Web sites, pages with link collections, pages of bulletin boards, pages with a personal diary and pages displaying some data in a table. tf-idf is a powerful method for documents consisting of many sentences such as news articles because it weighs keywords based on their frequency in the document. The pages browsed by User A are all news articles and include many sentences. Therefore the keyword precision and keyword recall of tf-idf has become high in those pages. The pages browsed by User B-E did not always include many sentences. Therefore the keyword precision and keyword recall of tf-idf has become low in those pages. Meanwhile the mouse-based method extracts keywords based on the user's mouse operation and does not consider the keywords' frequency in the documents. Therefore it extracts the keywords that the user was interested in from the browsed pages, even if those pages do not include many sentences. This shows that the mouse-based method can extract the keywords that the user was interested in at high accuracy even in various kinds of pages where tf-idf cannot achieve its best performance.

5.4 Conclusion of the Experiment

We confirmed that the target text part of text tracing, link pointing, link clicking and text selection operation includes keywords the user was interested in at higher accuracy than the whole text of the page. When we used these all four kinds of operations for keyword extraction in the mouse-based method, it could extract keywords that the user was interested in at high accuracy even for pages with miscel-

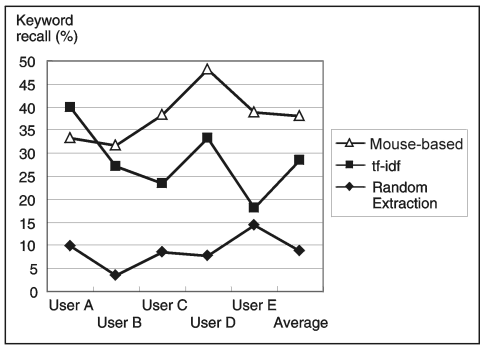

Figure 7: Keyword recall.

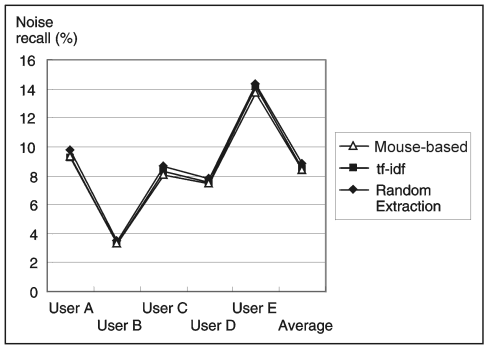

Figure 8: Noise recall.

laneous styles where tf-idf cannot achieve its best performance. Therefore we can expect a better Web page search by using keywords extracted by the mouse-based method for relevance feedback.

In the experiment, five users browsed their favorite Web pages as they usually do. The result shows that the mouse-based method extracts keywords that the user was interested in at high accuracy just by using the user's usual mouse operations. From this, we confirmed that the system can acquire the information about the user's interest without insisting users to answer the questionnaire about their interest or to grade the pages they have browsed. Therefore we can expect that the mouse-based method allows users to use the advanced functions for Web page search more easily without inputting keywords or rating pages consciously.

6. DISCUSSION

This section discusses the type of operation and the expiry of the user profile.

6.1 Type of Operation

This subsection will compare the four kinds of operations. Table 5 shows the keyword precision, the keyword recall and the noise recall in every type of operation considering all users. The keyword precision of operations performed unconsciously such as text tracing and link pointing is lower than that of operations performed consciously such as text selection and link clicking. There is difference in the keyword recall according to the type of operation. We can see that the keyword recall is not going to be high by using only operations with high keyword precision such as text selection and link clicking.

Table 5: Difference among the types of operations.

Type of operation	Keyword precision(%)	Keyword recall(%)	noise recall(%)
Text tracing	5.83	20.91	5.21
Link pointing	10.15	8.85	1.21
Link clicking	22.76	15.01	0.79
Text selection	50.00	1.34	0.02

When the system searches Web pages using a search query with a few keywords, operations with high keyword precision such as text selection and link clicking will be effective. When the system searches Web pages using vector space models with many keywords, operations with high keyword recall such as text tracing and link pointing will also be effective. In this case, the system can also change the weights of the keywords according to the type of operation used for the keyword extraction. It will be important to select the type of operation for the keyword extraction and to weighs keywords based on the type of operation according to the target application.

6.2 Expiry of User Profile

Some researchers work on the expiry of the user profile. Miyahara hypothesizes that the strength of the user's interest follows the Gamma distribution and tries to prove its correctness[21]. NewsT uses the genetic algorithm and leaves the genes with the current and strong interest[22]. SIFTER considers the history of relevance feedback (interesting or not on a category) as a Bernoulli trial and judges the change of occurrence probability by Bayesian analysis[23]. Crabtree categorizes browsed documents into fixed categories for a period of time and sees the differences between some periods[24]. IndexNavigator tries to infer the change of the user's interest by the hypothetical inference[25].

However these studies consider the change of the user's interest in a long term. This is because they are based on the user's interest in a page not in a text part of the page. Our mouse-based method selects the text part in a page. This is the biggest difference when comparing with other methods. The shortcomings is that our method just selects the text part in a page but does not consider the time while the selected keywords are valid. In the case of relevance feedback from the current page, our method is enough. However when the user wants to do relevance feedback based on the context of the current session, we need to consider the expiry or weights (which are changed during the session) of selected keywords.

7. CONCLUSIONS

This paper describes a method for extracting a text part which the user might be interested in using the user's mouse operation performed during his/her usual Web browsing. In our research, we conducted a preliminary survey and discovered four kinds of operations related to the users' interests: text tracing, link pointing, link clicking and text selection. We developed a system which extracts the target text part of these four kinds of operations by sentence or line. We conducted an experiment to see if the extracted text by the

mouse-based method is actually the part the user was interested in.

The result shows that the target text parts of every four kinds of operations include keywords the user was interested in at higher ratio than whole text of the page. Comparing the mouse-based method with the method for extracting keywords at random, we confirmed that the mouse-based method extracts keywords that the user was interested in at about 4 times of accuracy. These results shows that the mouse-based method extracts the text part that the user was interested in without insisting users to answer questionnaires.

We also compared the mouse-based method with tf-idf which is the most popular keyword selection method. The result shows that the mouse-based method extracts the keyword that the user was interested in at about 1.4 times of accuracy. The result also showed that the mouse-based method extracts keywords at high accuracy even for pages with miscellaneous styles such as bulletin boards and link collections where tf-idf does not achieve its best performance. Therefore we can expect a more sophisticated information retrieval using the extracted text by the mouse-based method for relevance feedback. Our future research will conduct a relevance feedback by using the keywords extracted by the mouse-based method and see its performance.

8. REFERENCES

[1] Broder, A., et al.: Graph Structure in the Web, *Proc. of the Ninth International World Wide Web Conference, In Computer Networks and ISDN Systems*, Vol. 33, pp. 309–320 (2000).

[2] Salton, G.: Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer, Addison Wesley (1989).

[3] Resnick, P., et al.: Recommender Systems, *Comm. of the ACM*, Vol. 40, No. 3, pp. 56–89 (1997).

[4] Belkin, N. J. and Croftk, W. B.: Information Filtering and Information Retrieval: Two Sides of the Same Coin?, *Comm. of the ACM*, Vol. 35, No. 12, pp. 29–38 (1992).

[5] Sugimoto, M.: User Modeling and Adaptive Interaction in Information Gathering Systems, *Journal of Japanese Society for Artificial Intelligence*, Vol. 14, No. 1, pp. 25–32 (1999).

[6] Mulvenna, M.D., Anand, S.S., Buchner, A.G.: Personalization on the Net using Web Mining, *Comm. of the ACM*, Vol. 43, No. 8, pp. 123–125 (2000).

[7] Shardanand, U. and Maes, P.: Social Information Filtering: Algorithm for Automating 'Word of Mouth', *Proc. of CHI'95*, pp. 210–217 (1995).

[8] Yan, T. W. and Garcia-Molina, H.: SIFT - A Tool for Wide-Area Information Dissemination, *Proc. of 1995 USENIX Technical Conference*, pp. 177–186 (1995).

[9] Resnick, P., et al.: GroupLens : An Open Architecture for Collaborative Filtering of Netnews, *Proc. of CSCW'94*, pp. 175–186 (1994).

[10] Pazzani, M. and Billsus, D.: Learning and Revising User Profiles: the Identification of Interesting Web Sites, *Machine Learning*, Vol. 27, No. 3, pp. 313–331 (1997).

[11] Lang, K.: NewsWeeder: Learning to Filter NetNews, *Proc. of ICML'95*, pp. 331–339 (1994).

[12] Smyth, B. and Cotter, P.: A Personalized Television Listings Service, *Comm. of the ACM*, Vol. 43, No. 8, pp. 107–111 (2000).

[13] Kantor, P.B., et. al: Capturing Human Intelligence in the Net, *Comm. of the ACM*, Vol. 43, No. 8, pp. 112–115 (2000).

[14] Morita, M. and Shinoda, Y.: Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, *Proc. of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, pp. 272–281 (1994).

[15] Sakagami, H. and Kamba, T.: Learning Personal Preferences on Online Newspaper Articles from User Behaviors, *Proc. of the Sixth International World Wide Web Conference, In Computer Networks and ISDN Systems*, Vol. 29, pp. 1447–1456 (1997).

[16] Yoshida, M. and Yoshitaka, A.: Digital Reminder: Building and Accessing a Real-World-Oriented Database, *Proc. of The 8th Workshop on Interactive Systems and Software (WISS'2000)*, pp. 101–110 (2000).

[17] Ohno, T.: IMPACT: Eye Mark Reusing Technique to Support Information Browsing Task, *Proc. of The 8th Workshop on Interactive Systems and Software (WISS'2000)*, pp. 137–146 (2000).

[18] Okumura, M.: Introduction of Natural Language Processing Tools, *IPSJ Magazine*, Vol. 41, No. 11, pp. 1203–1207 (2000).

[19] http://www.w3.org/DOM/

[20] Goodman, D.: Dynamic HTML The Definitive Reference, O'Reilly (1998).

[21] Miyahara, K. and Okamoto, T.: Quantified Estimation Method of User's Information Interests based on the Web Browsing and its Application to Collaborative Filtering, *Proc. of the Technical Report of IEICE*, ET97-115, pp. 17–24 (1998).

[22] Shetch, B. and Maes, P.: Evolving Agents for Personalized Information Filtering, *Proc. of IEEE Conference on Artificial Intelligence for Applications*, pp. 345–352 (1993).

[23] Mostafa, J., et al.: A Multilevel Approach to Intelligent Information Filtering: Model, System, and Evaluation, *ACM Transactions of Information Systems*, Vol. 15, No. 4, pp. 368–399 (1997).

[24] Crabtree, I.B. and Soltysiak, S.J.: Identifying and Trackign Changing Interests, *Iternational Journal of Digital Library*, Vol. 4, pp. 38–53 (1998).

[25] Osawa, Y., Sugawa, A. and Yachida, M.: An Index Navigator:Underatanding and Expressing User's Changing Interest, *Journal of Japanese Society for Artificial Intelligence*, Vol. 13, No. 3, pp. 461–469 (1998).