BEHAVIOR BASED WEB PAGE EVALUATION

GANESAN VELAYATHAN

The Graduate University for Advanced Studies

National Institute of Informatics, JAPAN

ganesan@grad.nii.ac.jp

SEIJI YAMADA

National Institute of Informatics, JAPAN

seiji@nii.ac.jp

This paper describes our efforts to investigate factors in user browsing behavior to automatically evaluate Web pages that the user shows interest in. To evaluate Web pages automatically, we developed a client-side logging/analyzing tool: the GINIS Framework. We do not focus on just clicking, scrolling, navigation, or duration of visit alone, but we propose integrating these patterns of interaction to recognize and evaluate user response to a given Web page. Unlike most previous Web studies analyzing access through proxies or servers, this work focuses primarily on client-side user behavior using a customized Web browser. First, GINIS unobtrusively gathers logs of user behavior through the user's natural interaction with the Web browser. Then, it analyses the logs and extracts effective rules to evaluate Web pages using a C4.5 machine learning system. Eventually, GINIS becomes able to automatically evaluate Web pages using these learned rules, after which the evaluation can be utilized for a variety of user profiling applications. We successfully confirmed, for example, that time spent on a Web page is not the most important factor in predicting interest from behavior, which conflicts with the findings of most previous studies.

1. Introduction

Over the last decade, the Web has grown substantially in size, popularity, scope and number of users [1, 2]. Many proposals have been made for tools and systems to assist with everyday user browsing and searching task [3, 4, 5, 6, 7, 8, 9, 10]. In recent years, systems enabling Web personalization for individuals have gained particular attention by researchers and commercial companies [34, 35, 41, 42, 43]. White and Drucker [49] investigated the use of user behavior for better understanding of searching task for personalization using client side logging. Hu et al [50] used user's browsing behaviors to predict demographic information like age and gender. This kind of information then used to personalize information for the user.

Web personalization can be described as actions "that make the Web experience of a user personalized to the user's taste" [37]. Most previous research and commercial tools for Web personalization have relied upon overt methods of asking users for their answers in order to construct user profiles [40, 41, 42, 43]. Building a user profile that adapts to a user's daily interests is a challenging task. This is because it is hard to predict which Web sites most interest the user without asking the user to interact explicitly with the system.

The reason conventional profiles are not adaptive is that the user evaluation process has been manual; that is, a user must interact with the system explicitly and tell the system each time a webpage is relevant to the task or not [41, 42, 43, 46, 47, 48]. This is laborious; it requires time and inclination, and users often forget to notify the system. A less intrusive approach to the construction of user profiles is required.

With this in mind, this study proposes a new method of automatically discovering and judging user interest based on user browsing behavior. User behavior is defined here as the habitual actions performed by users when browsing and searching, such as clicking links, bookmarking, printing or selecting text, and so on. When users use a Web browser to browse the Internet, they (whether intentionally or not) interact with the browser's interface [31]. If we can detect and learn these user's patterns of behavior, we can use these tendencies in individual browsing habits to perform automatic page evaluation. The behavior performed whenever a user browses and the patterns learned from previous users constitute a learned behavioral database. We believe that through comparison with this database, new websites visited can be automatically evaluated without placing any burden on the user. We are also hypothesizing that user browsing strategies and habits do not change greatly. By using this method, this study verifies that a highly accurate automatic self-constructed user profile can be created. Furthermore, we believe that browsing behavior is unlikely to change even if a user is browsing in a different language environment, which means a language-independent method of evaluating Web pages can be constructed. This study is similar to the work done by Fox et al [28]; however, the scope of their work was restricted to search engine result pages and subsequent clickthroughs.

This study also focuses on highly accurate logging at the client-side. Logging at the client-side enables a more complete grasp of user behavior than proxy sites or server sites. However, because logging user behavior at the client-side is a challenging task for which we needed to build high-level client software, we encountered various difficulties. Two issues emerged with regard to acquiring correct user browsing behavior. The first was that overt collection of data for experimental purposes was not desirable; in short, if users were conscious of an experimental environment, this would interfere with their usual browsing habits. The other issue is that it was necessary for the experimental instrument for data collection to be identical to the browser regularly used by the users. The approach of building a customized browser is similar to Claypool's Curious Browser [20]; however, we focus on keeping the browser as identical as possible to Internet Explorer.

Starting from these considerations, we undertook construction of the GINIS Framework, which expands the Internet Explorer Component available on .NET Framework from Microsoft. We used the GINIS Framework to conduct experiments with test subjects. This paper presents those findings.

The paper is organized into 8 sections. Section 2 describes problematic areas from previous and related research. Section 3 explains the details of Web page evaluation based on user behavior as well

as the GINIS Framework we have developed for client-side logging of user behavior. Section 4 describes the experiments conducted using the GINIS Framework and in Section 5, we present the result of these experiments. Section 6 discusses the usability of the GINIS browser, the data cleaning process, the classification rules, and the accuracy rates. Furthermore, we also present some current limitations of this research. Section 7 discusses future directions for this research, and Section 8 concludes the paper.

2. Current Practice and Research

2.1. Web Navigation

Interest in Web navigation has been an issue for commercial companies and researchers all around the world since the beginning of the WW

W. In the field of Web navigation, many previous studies are based on data found in server logs for analyzing various aspects of user navigations [26, 27]. However, data available in server logs only shows user behavior for a single site, and it is difficult to collect and aggregate all the logs to discover common rules of user behavior. Shahabi and Chen [44] have pointed out that Web server logs might be inaccurate because Web usage data from the server side are not reliable.

Little research has been done relating to client side web navigation. One of the latest studies is the work presented by Weinreich et al., [15] who observed 25 users for a mean usage duration of 105 days. Weinreich et al. focused on three aspects of Web navigation: changes in the distribution of navigation action, speed of navigation, and within-page navigation. Prior to this was the work presented by Catledge and Pitkow [11], (the first long-term client-side usage study made), observing 107 users over a duration of 21 days. Catledge and Pitkow identified different navigation strategies and found that users operate only on small areas within sites.

Another solution other than log analysis is using eye-tracking equipment to record users' visual attention [45]. Unfortunately, eye-tracking methods are relatively expensive, difficult to set up and administer, and the equipment can only be used with one person at a time. These methods are also limited to the laboratory, and cannot easily replicate the user's home or work environment.

Additionally, most of these previous studies tended to focus on tracking user behavior itself rather than looking for connections between user behavior and user interest.

Past works on user interest and Web usage behavior tended to focus more on navigation history, explicitly clicked links and the time spent on a Web site. Morita et al., [12] found that there is a strong tendency for users to spend a greater length of time reading articles of interest to that user, and to spend less time on pages of less interest. However, Weinreich [15] found that users spend less than 12 seconds on nearly 50% of the Web pages shown to them. This means that users make nearly 50% of their decisions to navigate to the next page before thoroughly reading the contents.

2.2. Personalization

Personalization is the process of presenting the right information to the right user at the right moment [16, 36]. In most cases, personalization begins with building a user profile based on user interest. Interest data can be collected either explicitly by asking for feedback from the user

regarding preferences or implicitly by observing user behavior, such as time spent reading a Web page.

Most research on personalization tend to focus on building user profiles based on requiring users either to fill out written or online questionnaires, or to specify Web page categories of interest [40, 41, 42, 43]. These conventional pre-defined user profiles must be manually constructed when a tool is initially used, and can be refined only through manual adjustment. However, since user interests change with time and task-orientation, every time a user's interest changes, the user has to manually re-create the user profile. Users may provide inconsistent or incorrect information, and the profile built up is static, whereas user interest may change over time, meaning that the construction of such a profile might place a burden on users that they do not wish to accept [16].

Research such as Chaffee's [18] proposes building user preferences automatically by studying user browsing history. Studies such as [16] focus on automatically learning from user search history. Many other studies, such as [19, 20, 38, 39] have focused on building accurate user profiles based on implicit user behavior.

To achieve effective personalization, profiles should distinguish between long-term and short-term interests, and should include a model of the user's context, i.e., the task in which the user is currently engaged and the environment in which they are situated [16, 17].

2.3. Implicit Feedback

Implicit feedback techniques unobtrusively obtain information about users by observing their natural interactions with the system [21, 25, 29, 30]. The user behaviors most extensively investigated as sources of implicit feedback include reading time, as well as saving, printing and selection behaviors. The primary advantage to using implicit techniques is that such techniques remove the user's burden of from providing explicit feedbacks [21, 25].

Most previous implicit feedback research experiments have been conducted in a controlled laboratory and have tended to focus on time spent on a Web site as one of the most important factors [12, 22, 23, 25]. Morita et al., [12] investigated reading time as a feedback factor, Seo and Zhang considered user browsing patterns [24], and Kelly and Belkin [30] investigated the use of reading time and scrolling as an indication of relevance. Claypool et al., [20] identified several implicit measures such as time spent viewing a page and the amount of scrolling on a page, suggesting that the combination of time and scrolling had a strong positive correlation with the explicit ratings. Furthermore, Claypool also concluded that the number of mouse clicks and individual scrolling measures were found to be ineffective in predicting the explicit ratings. Fox et al., [28] found that dwell time, position, scroll count and exit type are predictive actions of relevance judgments for individual Web pages and that dwell time, number of results listings and exit type are more predictive of overall session satisfaction.

However, there are many other factors to consider. This point remains in question, continuing to motivate research towards its resolution.

3. Research Approach

When browsing or searching the Web, we tend (intentionally or not) to interact with the browser's interface. For example, we print, save, bookmark, move the mouse, click a link, and so on. This study will refer collectively to these interactions as *navigation actions*. Behavior such as printing, bookmarking, saving page, etc. is termed *direct behavior*. Kim [17] found that 95% of users are interested in the contents when they perform these kinds of direct actions. However, [15] found that only 2% of all user activity constitutes this direct behavior. From this we can conclude that users will not necessarily create bookmarks or print even on those pages in which they have indicated interest. These things are known from experience. Profiles constructed by taking direct behavior as the only target can be expected to produce results to a certain extent, but as long as users are not actively undertaking this direct behavior, discovering any generalized rules is extremely difficult.

In order to build user profiles automatically, we believe that implicit user behavior (*indirect user behavior*) will play a major role. Careful study of indirect user behavior could help us to better understand user interest, and thus to further enhance techniques for Web personalization.

3.1. Behavior Attributes

A preliminary study was first made, monitoring one user's browsing behavior using video and screen shot captures. After carefully studying these screen shots and taking into consideration the technical programming restrictions, we chose around 40 specific user behaviors and built a monitor for the browser. The details of behaviors monitored are described below. Some less important behaviors, such as Windows Update and View Help, among others, were omitted during the programming stage, as we considered these actions less significant for the study of user behavior. The behaviors were selected using a pilot monitoring system, available documents of .Net 2.0 Browser Objects, and video/screen shot capture monitoring.

We use the term navigation action here to describe the individual "components" of user behavior in performing actions using the browser directly. We use the term user behavior to describe the result achieved by performing these navigation actions. For example, the navigation actions "Hit Backspace Key," "Click Back Arrow on Menu," or "Use Back Button on Mouse" all constitute the same "Move Back" user behavior.

Some user behaviors were gathered at the navigation action level rather than at the user behavior level; this is because we preferred to keep the granularity of the log as high as possible. During the experimental stage, we gathered this information at the navigation action level, after which we pre-processed the log and construed user behavior. So, for example, if a user performed "Click Location" and "Drag Mouse while Holding Left Mouse Button" navigation actions, this would be collectively recorded as "Highlight Text."

Over 70 navigation actions and around 40 user behaviors were logged during this experiment. The logging attributes were designed based on the work presented by Catledge and Pitkow [11], with some additions. Outdated attributes were left out during the system design stage. Table 1 shows a short sample of user behaviors added.

User Behavior	Description of Behavior / Action
Bookmark to Desktop	Bookmark to Desktop
Change Encoding	Change Encoding
Change Font Size	Change Font Size
Close Current Tab	Close Current Tab
Open Favorite Bar	Open Favorite Bar
Open Link to New Tab	Open New Page in Tab Window
Open New Tab	Open New Tab
Open Search Bar	Open Search Bar
Overwrite To File	Overwrite the Existing File
Show Full Screen	Show in Full Screen
Text Copy	Copy the Text
Text Copy All	Copy All the Text
Text Cut	Cut the Text
Text Highlight	Highlight the Text
Text Paste	Paste the Text
Resize Browser	Resize Current Browser

Table 1 Sample of User Behaviors Added

3.2. GINIS Framework Architecture

A framework called GINIS was developed during this research for the purpose of logging user behavior at the client-side. The GINIS Framework was built based on .NET Framework 2.0, using Microsoft Internet Explorer Browser Component. It is an extendable framework intended for client-side logging.

The experiment for collecting user behavior was divided into 2 separate parts. The first part was the learning part, during which the users were prompted to answer every time before navigating to a new page whether they liked the page they just viewed. This favorable information (along with behavior data) was collected and stored in a database.

The learning engine refined the logs from the raw data database and stored it in the User Behavioral Database (UBD). Classification learning (building a decision tree using C4.5) was performed based on the information form the UBD. During the testing stage, the prediction engine generated and compared the user's new behavior with the behavioral patterns stored in the UBD, and the user's implicit interest in a particular page was predicted based on this comparison.

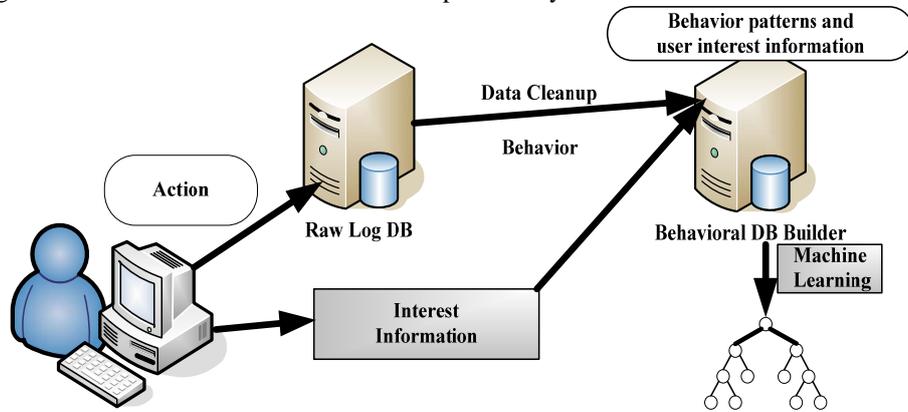Figure 1 and 2 show an overview of the 2 steps of the system built for this research.
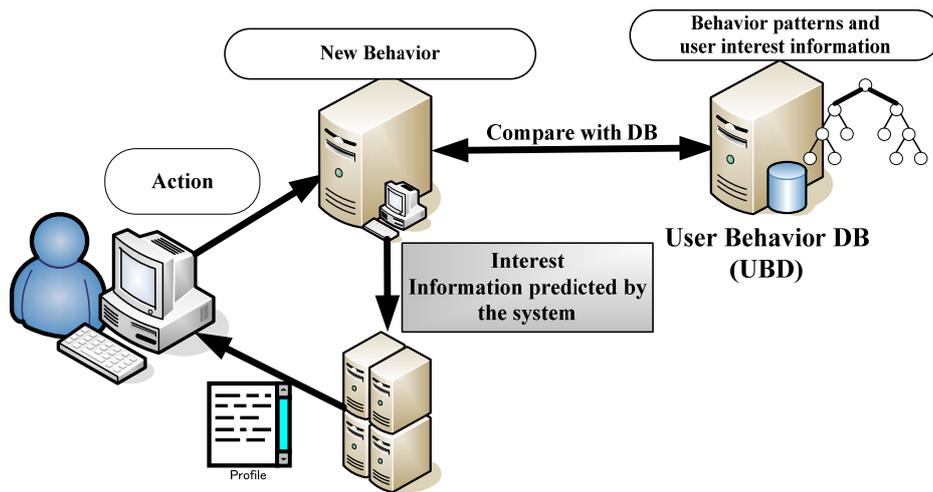

Figure 1 GINIS Learning Stage

Figure 2 GINIS Testing Stage

The GINIS Framework consists of 4 main modules: a client interface to detect and log user behavior (the browser), a database to store the user log information (the raw logger), a learning engine and a prediction engine.
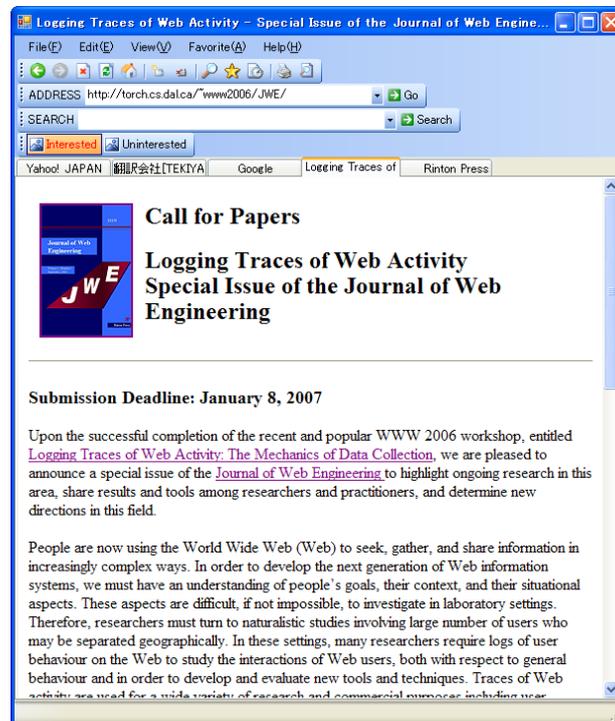
Figure 3 GINIS Browser User Interface

The browser: Used to log user navigation actions, this mimics almost all of Internet Explorer 6, with some extension of the right-click menu. A simple AJAX plug-in and form input plug-in was added to enhance the browsing and logging features for Japanese IME keystroke detection. Naturally, it is a simple tab browser; because of the popularity of tab browsers in the Japanese community. Since Internet Explorer 7 will be tab-browsable, we hope that the result of this study will be useful in the future. The user interface menu of GINIS browser is shown in Figure 3.

The logger: A real-time user navigation logger, this logs user navigation behavior, including primitive navigation actions like mouse movements, scrolls, clicks, bookmarking, etc., as well as more advanced combinations of navigation actions like Text Highlight, Text Copy etc.

The analyzer: A behavioral database builder, this builds a decision tree based on user interest information. The C4.5 classification learning system, a landmark decision tree machine learning method widely used in practice was chosen as the classification algorithm [13]. Survey [14] show that it provides good classification accuracy and it is the fastest among the compared *main-memory* classification algorithms for machine learning and data mining. Another reason C4.5 was chosen in preference to other algorithms its ability to generate rules from the decision tree. C4.5 rules are shown as linguistic information (such as *if-then* type rules) which people can easily comprehend and use.

The predictor: Based on the classification learning built at the learning stage using the analyzer, the predictor module will automatically classify and assign user behavior to "*of interest*" and "*not-of-interest*" classes. Based on the user behavior performed, the predictor module also has the ability to classify Web pages viewed by a user as "*pages of interest*" and "*pages of non-interest,*" which can be used to build profiles for personalization.

4. Experiments

4.1. Overview

We conducted systematic experiments using the GINIS Framework to gather data in order to discover new rules linking user behavior and interest. Discovered from the data gathered from the subjects, the rules were summarized either as shared user behavior rules and individual user behavior rules. C4.5 was used as the classification algorithm. The C4.5 system can deal with both numerical information and character string information at the same time, and it can represent a decision tree with a form of *if-then* type rules, making it more effective than other tree-like classification methods. Verification was performed using the 10-fold validation method. We designed the experiment to be performed within an open environment in order to expose and clarify unintentional user behavior while browsing (one of the main objectives of the study). Each user installed the GINIS browser on his or her desktop and used it in the conduct of his or her daily activities. We did not control or restrict the user in any way, preferring to keep the experiment as non-laborious as possible.

When using the GINIS browser, a questionnaire dialog box pops up every time the user undertakes a "Next Page" navigation behavior (by using any navigation action to move from a page to a new page). On this questionnaire, users could choose "***like,***" "***dislike,***" or "***unknown***" in regard to the page indicated. The default answer was set to "unknown," as users tend to click without really considering the question when they are very busy with deadlines etc. The browser also allows the user to terminate the popup, instead using the question button at the side of the menu bar. Before the experiment, we instructed the participants to choose the pages that interested them based on their experience on Web surfing.

The log file generated by the GINIS Framework contains data on the date, time, Window ID, Tab ID on which the behavior occurred, the circumstances surrounding how the behavior took place (use of toolbar, use of menu, use of right click, mouse, keyboard, short cuts etc.), what the behavior represented (printing out a page, copying text, etc.), the URL of the page where the event occurred, and the user interest information (whether the page was ***liked***, ***disliked***, or ***unknown***). Furthermore, Web page text data, detail query keywords, and character strings inputted via keyboard were collected for future use.

4.2. Data Gathering

Ten unpaid volunteers (6 male and 4 female) were recruited to participate in this study. They ranged in age from 21 years old to 38 years old (mean age: 29.1 years). All of them had between 4 and 12 years of Web browsing history (mean history: 8.3 years). Almost all the participants used the internet for more than 6 hours per day (mean usage time: 10 hours per day), mainly either during work hours or for pleasure and entertainment. All participants were bilingual (Japanese and English);

3 worked as translators, 3 as software engineers, 2 as office workers, and 2 were fulltime students. The subjects used the GINIS browser for a period of 16 to 31 days (mean usage duration: 22 days).

4.3. Preliminary Data Processing

GINIS was built with very high logging functionality to enable future use of the data. The original log file comprises over 460,000 lines, including 195,000 action data. In the preliminary data processing stage, we performed a data clean-up to 65%. Around 65,000 lines were left after this process. Most of the removed data were mouse movement logging (mouse locus and axis) and text copied/pasted to/from the clipboard by the user. This text information was logged but has not yet been used. We believe that these logs could be used in future for building user profiles for personalization.

4.4. Data Summary

Table 2a and 2b, shows the 5 most and 5 least frequently performed behaviors. Behaviors that did not take place were omitted. The total number of behaviors logged from the users was 64,312. There was an average 5.2 user behaviors per page, with the total visited pages during this experiments being 12,368 pages.

Behavior	Frequency
Scroll	19091 times
Key Input	14188 times
Form Input	9329 times
Navigation Link	4585 times
Search Text	1284 times

Table 2a: Occurrence of the 5 most frequent behaviors

Behavior	Frequency
Go Forward	126 times
Stop Loading	88 times
Add to Favorite	79 times
Print	64 times
Save As	2 times

Table 2b: Occurrence of the 5 least frequent behaviors

On-page "Scroll" was the most common behavior performed on a page. We do support the finding of [15] that "Navigation Link" is still the most important behavior performed on the browser. Even though our results showed "Navigation Link" occurred only 4585 times, placing it fourth, the

three behaviors which were ranked higher in occurrence (Scroll, Key Input, and Form Input) all take place multiple times while "Navigation Link" tends to be performed singly.

5. Results

5.1. Result A: Discovery of Shared User Behavior Rules

5.1.1. Overview

Classification learning was performed using the C4.5 machine learning algorithm. 2856 cases were used as training data. Here, a case means a set of user behavior or navigation action data and the evaluation of the Web page when such behaviors take place. The C4.5 error rate was set at 25%. As a result, 13 rules were obtained. Eight of these were found to be rules governing pages "*not of interest*", and 5 of these were rules governing pages "*of interest.*" Out of 2856 training data, 1997 pages were classified as "*of interest*" and 859 pages were classified as "*not of interest*" by the test subjects.

2027 pages (70.97%) were correctly classified, and 829 pages (29.03%) were incorrectly classified at this point. As a result of performing 10-fold cross validation, the Confusion Matrix indicated in the following table was obtained. Table 3 shows the summary of classification learning results before removing instances of inconsistency.

User Evaluation	Classified as "*of interest*"	Classified as "*not-of-interest*"	Total
Of Interest	1852	145	1997
Not of Interest	684	175	859

Table 3 Classification Learning Results (before removing the instances of inconsistency)

As a result of validating the data, we found that our evaluation data contained many instances of inconsistency. Here, the term inconsistency means that the cases of training data have the same type of behavior occurrence but different kind of user evaluation. For example, if we had three cases of training data where all the occurrence of user behaviors are identical but the evaluations of two of the cases are evaluated as "*of interest*" and one is evaluated as "*not of interest*", we remove the "*not of interest*" from the training data, and used only the remaining two cases of "*of interest*". Furthermore, when we only have identical number of cases, for example, if one was evaluated as "*of interest*" and the other one was evaluated as "*not of interest*", we randomly choose the case to be included as the training data. Following on from this, we performed further clean-up of the data.

As a result, 2249 cases were left, and within these 1885 were judged as "*of interest*" by the user and 364 were judged as "*not of interest*" by the user. Of these evaluations, 2005 pages (89.15%) were correctly classified, and 244 pages (10.85%) were incorrectly classified. This represented a significant improvement in classification capability. Table 4 shows the summary of classification learning results after removing the instances of inconsistency. Furthermore, Table 5 shows the summary of Result A.

User Evaluation	Classified as "*of interest*"	Classified as "*not-of-interest*"	Total
Of Interest	1808	77	1885
Not of Interest	167	197	364

Table 4 Classification Learning Results (after removing the instances of inconsistency.

Item	Before	After
Total Number of Instances	2856	2249
Correctly Classified Instances	2027 (70.97%)	2005 (89.15%)
Incorrectly Classified Instances	829 (29.03%)	244 (10.85%)
Number of "*of interest*" rules	5	5
Number of "*not of interest*" rules	8	21
Mean Absolute Error	0.158	0.139
Root Mean Squared Error	0.385	0.296
Relative Absolute Error	91.61%	51.07%
Root Relative Squared Error	99.36%	80.38%
Precision of "*of interest*"	0.730	0.915
Recall of "*of interest*"	0.927	0.959
Precision of "*not of interest*"	0.547	0.719
Recall of "*not of interest*"	0.204	0.541

Table 5 Summary of Result A

5.1.2. Rules

Rules were created using C4.5, after first removing inconsistent evaluations. Twenty-six rules were output in total. 5 of these were found to be rules governing "*of interest*", and 21 of these were rules governing "*not of interest*." All the rules "*of interest*" and the "*not of interest*" are presented below. The default class was set to "*of interest*". The term "Stay Time" in these rules is indicated in seconds; all other terms are indicated by number of occurrences of that particular user behavior. Stay Time is the total time on a page in seconds that the browser window was active (i.e., the window was set on focus mode). We assume that when the browser window is active, the user is reading or interacting with the browser. If the browser window is not active, the time calculation will stop. The term "Key Input" represents all keyboard stroke actions, while the term "Form Input" represents user behavior in inputting for forms available on Web pages (using the keyboard).

5.1.3. Rules "Of Interest"

Figure 4 indicates the 5 rules for pages tagged "*of interest*" by C4.5 classification system.

Rule 112:
  Scroll <= 1
  Form Input > 1
  Key Input <= 2
  Move Back <= 0
  → **Class Of Interest [97.5%]**

Rule 67:
  Scroll <= 3
  Reload <= 0
  Search Text <= 0
  Form Input <= 3
  Move Back <= 0
  → **Class Of Interest [93.5%]**
Rule 101:
  Scroll > 3
  Reload <= 0
  Form Input > 1
  Navigate <= 0
  Key Input <= 8
  Move Back > 0
  → **Class Of Interest [84.5%]**

Rule 15:
  Scroll <= 0
  Reload <= 0
  Key Input <= 6
  → **Class Of Interest [94.3%]**

Rule 25:
  Reload <= 0
  Form Input >= 1
  → **Class Of Interest [87.2%]**

Figure 4 Rules tagged as "of interest"

5.1.4. Rules "Not of Interest"

Figure 5 indicates the 21 rules for pages tagged "*not of interest*" by C4.5 classification system.

Rule 2:
  Scroll <= 0
  Search Text <= 0
  Form Input <= 0
  Key Input > 0
  Key Input <= 5
  Move Back <= 0
  Text Copy <= 0
  → **Class Not of Interest [92.6%]**
Rule 28:
  Scroll <= 0
  Key Input > 9
  Key Input <= 10
  Move Back <= 0
  Text Copy <= 0
  → **Class Not of Interest [84.1%]**
Rule 20:
  Form Input > 2
  Move Back > 1
  → **Class Not of Interest [82.0%]**

Rule 69:
  Stay Time > 3
  Scroll > 0
  Search Text <= 0
  Form Input > 3
  Form Input <= 4
  → **Class Not of Interest [88.2%]**

Rule 57:
  Search Text <= 0
  Form Input <= 0
  Key Input <= 0
  Move Back > 0
  → **Class Not of Interest [83.7%]**

Rule 52:
  Stay Time > 3
  Stay Time <= 5
  Scroll > 0
  Scroll <= 16
  Form Input > 0

Rule 24:
   Stay Time <= 16
   Scroll <= 0
   Form Input > 0
   Form Input <= 1
   Key Input > 6
   Key Input <= 35
   → **Class Not of Interest [80.6%]**

Rule 117:
   Scroll > 1
   Reload > 0
   Form Input > 1
   → **Class Not of Interest [77.0%]**

Rule 21:
   Scroll <= 0
   Key Input > 2
   Key Input <= 6
   Move Back > 0
   → **Class Not of Interest [75.8%]**

Rule 116:
   Reload > 0
   Form Input > 2
   Move Back > 0
   → **Class Not of Interest [75.8%]**

Rule 34:
   Form Input > 2
   Key Input > 6
   Key Input <= 7
   → **Class Not of Interest [70.7%]**

Rule 78:
   Scroll > 1
   Scroll <= 2
   Search Text > 0
   Form Input > 1
   Form Input <= 3
   Key Input <= 1
   → **Class Not of Interest [66.2%]**

Rule 103:
   Navigate Link > 0
   Move Back > 0
   → **Class Not of Interest [63.0%]**

Rule 81:
   Scroll > 0
   Search Text > 0
   Form Input > 3
   → **Class Not of Interest [56.6%]**

   Form Input <= 1
   → **Class Not of Interest [82.0%]**

Rule 36:
   Stay Time > 128
   Stay Time <= 259
   → **Class Not of Interest [79.4%]**

Rule 108:
   Reload > 0
   Search Text <= 0
   Form Input <= 1
   Text Copy <= 0
   → **Class Not of Interest [75.9%]**

Rule 32:
   Form Input <= 3
   Key Input > 15
   Key Input <= 16
   → **Class Not of Interest [75.8%]**

Rule 17:
   Search Text > 0
   Form Input > 0
   Form Input <= 2
   Move Back > 0
   →**Class Not of Interest [70.7%]**

Rule 99:
   Stay Time > 5
   Scroll > 3
   Search Text <= 0
   Form Input > 1
   Move Back <= 0
   → **Class Not of Interest [68.6%]**

Rule 39:
   Stay Time > 20
   Reload <= 0
   Form Input > 3
   Key Input > 6
   → **Class Not of Interest [63.0%]**

Rule 83:
   Scroll > 3
   Form Input > 1
   Form Input <= 2
   Move Back <= 0
   → **Class Not of Interest [59.3%]**

Figure 5 Rules tagged as "not of interest"

5.1.5. Conclusion

After removing inconsistent evaluations, classification accuracy improved dramatically. Dramatic improvement can also be seen in the precision and recall data. We found that most of the rules generated were not governed by Stay Time, which conflicts with the findings of previous study [12, 20]. As Rule 36 shows (with an accuracy of 79.4%), when time spent on a page is in the range of 2-4 minutes, the Web page is likely to be "*not of interest*" to the user.

5.2. Result B: Discovery of Individual User Behavior Rules

5.2.1. Overview

It is natural to expect that browsing behavior will vary with the individual habits of each user. For Result B, the individual behavior of each user was analyzed and rules were generated. The experimental data used to generate Result B were the same as those used for Result A, listed above. The classification learning process was performed on a per-user basis after removing each instance of inconsistency. Figure 6 shows the percentage of correctly classified instances per participants. Table 6 shows the 2 users detail statistical data.
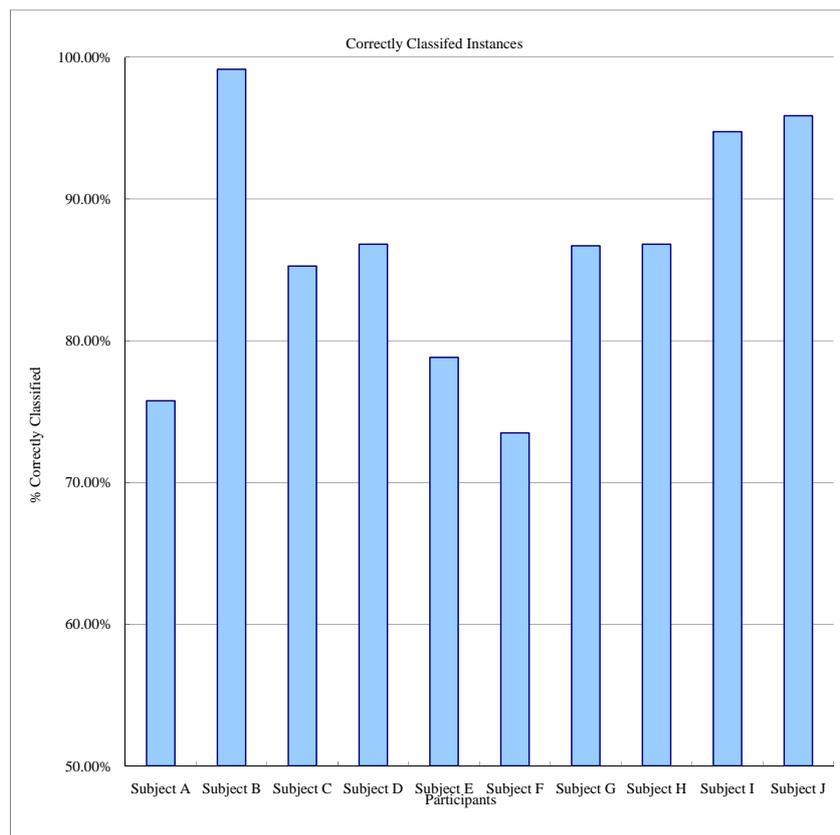

Figure 6 The percentage of correctly classified instances, by participants

Item	Subject A	Subject B
Total Number of Instances	297	353
Classified as "*of interest*"	163	342
Classified as "*not of interest*"	134	11
Correctly Classified Instances	225 (75.75%)	350 (99.15%)
Incorrectly Classified Instances	72 (24.24%)	3 (0.85%)
Number of "*of interest*" rules	4	2
Number of "*not of interest*" rules	6	2
Mean Absolute Error	0.3022	0.0168
Root Mean Squared Error	0.4492	0.0920
Relative Absolute Error	61.02%	26.59%
Root Relative Squared Error	90.27%	52.91%
Precision of "*of interest*"	0.731	0.991
Recall of "*of interest*"	0.883	1.000
Precision of "*not of interest*"	0.810	1.000
Recall of "*not of interest*"	0.604	0.727

Table 6 Summary of Subject A and B

5.2.2. Rules Generated for Subject A

For Subject A, 4 rules of interest and 6 rules of non-interest were generated by C4.5. The default rule was set to "*of interest.*"

The detailed rules for Subject A are shown in Figure 7.

```
Rule 31:                              Rule 7:
  Navigate Link > 0                     Scroll > 0
  Text Copy > 1                         Scroll <= 5
  → Class Of Interest [90.2%]           Form Input > 0
                                        Form Input <= 2
                                        → Class Of Interest [88.2%]
Rule 16:                              Rule 26:
  Stay Time <= 7                        Scroll <= 5
  Search Text > 0                       Search Text <= 0
  → Class Of Interest [84.3%]           Form Input > 1
                                        Key Input(Keyboard) > 8
                                        Key Input(Keyboard) <= 32
                                        → Class Of Interest [78.9%]
Rule 24:                              Rule 3:
  Scroll > 3                            Form Input > 0
  Form Input > 1                        → Class Of Interest [75.7%]
  → Class Of Interested [77.2%]
```

Rule 9:	Rule 6:
Scroll > 1	Scroll <= 0
Scroll <= 8	Form Input > 0
Form Input > 1	Form Input <= 2
Form Input <= 2	Navigate <= 0
Navigate <= 0	Key Input(Keyboard) <= 8
Key Input <= 1	→ **Class Not of Interest [90.3%]**
→ **Class Not of Interest [91.5%]**	
Rule 17:	**Rule 32:**
Search Text < 0	Navigate Link > 0
Form Input < 3	Move Back > 0
→ **Class Not of Interest [75.8%]**	→ **Class Not of Interest [50.0%]**

Figure 7 Rules generated for Subject A

For Subject A, Rule 31 shows that when navigating to a new page and performing the "Text Copy" behavior, the Web page viewed is 90.2% likely to be "*of interest*". Rule 24 and Rule 3, with accuracies of 77.2 % and 75.7% respectively, show that the Web page viewed is likely "*of interest*" when Subject A scrolls more than 3 times and either inputs some text on the page or inputs an online form.

5.2.3. Rules Generated for Subject B

For Subject B, a total of 4 rules, (2 "*of interest*" and 2 "*not of interest*") were generated by C4.5. The default rule was set to "*of interest*."

The detailed rules for Subject B are shown in Figure 8.

Rule 2:	Rule 8:
Scroll > 3	Reload > 0
Text Copy > 1	Search Text > 1
→ **Class Of Interest [97.3%]**	Form Input > 2
	→ **Class Interested [89.1%]**
Rule 3:	**Rule 13:**
Reload <= 0	Scroll < 3
Key Input <= 1	Key Input <= 1
→ **Class Not of Interest [66.2%]**	→**Class Not of Interest [33.3%]**

Figure 8 Rules generated for Subject B

5.2.4. Conclusion

The rules generated for Subject B had low accuracy for "*not of interest*" because there were generally very few instances classified as "*not of interest*" by Subject B. As for instances "*of interest*", Rule 2 predicts with an accuracy of 97.1% that when Subject B is viewing a page and has both scrolled more than 3 times and copied some text on the page, the Web page being viewed is "*of interest*."

6. Discussion

6.1. Browser Usability

After the experimental stage, we asked the 10 test subjects to evaluate the browser with 3 simple questions: How is the usability? Does GINIS contain all the features of a browser for normal daily use? Were you able to browse as if using your normal favorite browser? We obtained a mean score of 4.2, 4.3, & 4.3 for these questions respectively (on a 5-point evaluation scale). This demonstrates that the browser is easy to use and that users could navigate through their daily browsing without additional stress. A short survey was performed after the participants finished the experiments. One of our concerns was whether the process of rating affected user behavior; however, all the users notified us that since they had total control of the rating process (users were able to turn off the rating process whenever they were busy), they did not feel that the rating process affected their browsing behavior. However, it is impossible to judge this problem correctly at the moment.

We choose Internet Explorer over other browsers because it is easy to log high-granularity data using the provided Microsoft .Net Framework. Despite the efforts made by other browsers, Internet Explorer still maintains over 70% share of usage [32, 33]. It is the most commonly-used browser at the moment, making it easier to recruit subjects for experiments.

6.2. Data Cleaning

Two things that affected our clean-up result were Google's result list page and the use of online dictionaries. Most of the form input takes part on these pages. It is quite common for Japanese Web users to consult online dictionaries. All our participants used a few freely available online dictionaries, which gave us considerable difficulty in cleaning the results. The participants in this study tended to search for a particular word and then evaluate the answer of the query. In future experiments, we hope to restrict these sorts of pages from our results. These restrictions were not applied at the time of this initial experiment, so we expect more accurate results in the future.

6.3. Classification Rules

One early objective of this work was to locate common features of Web browsing shared by all users. After all user behaviors were gathered, the removal of inconsistent evaluations yielded a tremendous improvement in classification learning accuracy.

From Result A, mean accuracy was 91.4% (SD=4.80), precision was 0.915, and recall was 0.959 for the rules "*of interest,*" which was a very satisfying result for this classification learning.

For the classification "*not of interest*", mean accuracy for the 21 rules generated was 74.8% (SD=9.32), with precision of 0.719 and recall of 0.541. Over the experiment, only 364 pages were classified as "*not of interest*", which is around 18.2 % of the total data. Since the experiment was not performed in a controlled environment, it was difficult to collect an equal number of "*of interest*" and "*not of interest*" Web pages.

Most rules were found not to be governed by the time a user spends on a Web page. This conflicts with the findings of previous study [12, 20] but does support the findings of Kelly and Belkin [30]. We believe that this represents a major finding for this research.

6.4. Limitations of this Research

At present, our research was limited to searching for patterns of user behavior from all the behaviors available during browsing. We did not make detailed examination of the order of occurrence of the behaviors themselves. Also, we were not able to take more granular behaviors into consideration. For example, the user behavior "Scroll" might actually be broken into "Scroll Up", "Scroll Down", "Scroll Jump Up", "Scroll Jump Down," which might then be studied in further detail by taking into consideration the means and situation of occurrence of this behavior (whether mouse or keyboard was used, etc.).

One more factor limiting this research at the moment is that the log was collected in detail, but the log cleaning process was performed semi-automatically. This is because navigation actions tend to occur simultaneously, often making it difficult to determine which navigation action is part of which navigation behavior.

In regards to classification learning, no comparisons were made between a user's individual classification and the all-user aggregate classification learning. In order to investigate this problem, we need a much larger number of evaluated Web pages, in addition to the training data that we collected during the experiment. It will cost more to gather additional Web pages, and we expect that these experiments and comparisons will be done in the next stage of our work.

7. Ongoing and Future Work

The scroll action is the most frequently performed action on a page (after mouse action was removed from the log during our cleanup); we assume that more detailed study of this scroll action will result in a better understanding of action/behavior conversions.

A great number of mouse locus were logged during this experiment and these still have to be studied carefully. There is a strong possibility of finding new behaviors as well as valuable new research ideas.

At present, we have only taken into consideration the frequency of occurrence of a behavior on a single page during classification. There is a great need to carefully study the order of appearances, which we think would give us a better understanding of the relationship between user interest and user behavior.

We also intend to consider sessions of behavior, such as looking at the sequence of Web pages that led to a user's successful outcome, rather than just the final page of a successful outcome. For example, when a user performs form input (such as making online applications or purchasing items online), there is a strong possibility that a user was interested in any pages visited which led to or helped in the process of deciding to purchase a particular item or to input in a particular form.

One more interesting issue arising during this work is the need for a standard format for client-side logging, (e.g. a standard XML format), so that other researchers can collaborate and parse logs

easily and perform other methods of machine learning or data mining. As with many experiments, the greatest amount of time was spent collecting and cleaning up data.

Browsing behavior may well differ depending on whether the user is searching for something in particular or just using the Web for entertainment. This is an area where further research may be particularly useful.

8. Conclusions

We have presented here a method to automatically detect and log user behavior at the client-side by creating a client-side browser (the GINIS Framework). We used this framework to conduct an open environment field study, rather than a controlled laboratory study.

Past research has tended to focus on the time spent on a Web page as the main factor in evaluating implicit user interest. Our experiment confirms that the time spent on a page is not the only main factor; scroll action, form action, searching text, copying text etc. should also be taken into consideration as factors in evaluating user interest.

References

1. Gulli, A. and Signorini, A. (2005) The Indexable Web is more than 11.5 Billion Pages, In Proceedings of the WWW Conference 2005, 902-903.

2. Sullivan, D. (2005) Search Engine Sizes, http://searchenginewatch.com/showPage.html?page=2156481

3. Google Toolbar, http://toolbar.google.com/

4. Internet Explorer, http://www.microsoft.com/windows/products/winfamily/ie/default.mspx

5. Firefox, http://www.mozilla.com/

6. Opera, http://www.opera.com/

7. Nagino, N. and Yamada, S. (2003) Future View: Web Navigation Based on Learning User's Browsing Patterns, In Proceedings of the International Conference on Web Intelligence.

8. Kato, H., Nakayama, T. and Yamane, Y. (2000) Navigation Analysis Tool based on the Correlation between Contents Distribution and Access Patterns, In Proceedings of WebKDD Workshop on Web Mining for E-Commerce at the 6th ACM SIGKDD, 95-104.

9. Drummond, C., Ionescu, D., Holte, R., Georganas, N. and Petriu, E. (1996) Intelligent Browsing for Multimedia Applications, In Proceedings of International Conference on Multimedia Computing and Systems.

10. Milic-Frayling, N., Sommerer, R. and Rodden, K. (2003) WebScout: Support for Revisitation of Web Pages within a Navigation Session, In Proceedings of the International Conference on Web Intelligence.

11. Catledge, L. D. and Pitkow, J. E. (1995) Characterizing Browsing Strategies in the World-Wide Web, In Proceedings of the Third international World-Wide Web Conference on Technology, Tools and Applications, 1065-1073.

12. Morita, M. and Shinoda, Y. (1994) Information Filtering Based on User Behavior Analysis and Best Match Text Retrieval, In Proceedings of the 17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 272-281.

13. Quilan, R. (1993) C4.5: Programs for Machine Learning, Morgan Kaufmann.

14. Lim, T., Loh, W. and Shih, Y. (2000) A Comparison of Prediction Accuracy, Complexity, and Training Time of Thirty-Three Old and New Classification Algorithms, Machine Learning Journal, Vol. 40, 203-228.

15. Weinreich, H., Obendort, H., Herder, E. and Mayer, M. (2006) Off the Beaten Tracks: Exploring Three Aspects of Web Navigation, In Proceedings of the WWW Conference 2006, ACM Press, 133-142.

16. Speretta, M., and Gauch, S. (2004) Personalizing Search Based on User Search Histories, In Proceedings of the 13th International Conference on Information and Knowledge Management.

17. Kim, H., and Chan, K. (2005) Implicit Indicators for Interesting Web Pages, Web Information System and Technologies, 270-277.

18. Chaffee, J., Gauch, S. (2000) Personal Ontologies for Web Navigation, In Proceedings of the 9th International Conference on Information and Knowledge Management, 227-234.

19. Chen, C. C., Chen, M. C., and Sun, Y. (2001) PVA: A Self-Adaptive Personal View Agent, In Proceedings of the 7th International Conference on Knowledge Discovery and Data Mining, 257-262.

20. Claypool, M., Le, P., Waseda, M., and Brown, D. (2001) Implicit Interest Indicators, In Proceedings of the 6th International Conference on Intelligent User Interfaces, 33-40.

21. Kelly, D. and Teevan, J. (2003) Implicit Feedback for Inferring User Preference: A Bibliography, SIGIR Forum, 18-28.

22. White, R. W., Ruthven, I. and Jose, J. M. (2002) Finding Relevant Documents using Top Ranking Sentences: An Evaluation of Two Alternative Schemes, In Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 57-64.

23. Oard, D.W. and Kim, J. (2001) Modeling Information Content Using Observable Behavior, In Proceedings of the 64th Annual Meeting of American Society for Information Science and Technology.

24. Seo, Y. and Zhang, B. （2000）Learning User's Preferences by Analyzing Web-browsing Behaviors, In Proceedings of the Fourth International Conference on Autonomous Agents, 381-387.

25. Jansen, B. J. and M. D. McNeese (2005) Evaluating the Effectiveness of and Patterns of Interactions with Automated Searching Assistance, Journal of the American Society for Information Science and Technology, Vol. 56(14), 1480-1503.

26. Jansen, B. J., Spink, A., and Saracevic, T. (2000) Real Life, Real Users, and Real Needs: A Study and Analysis of User Queries on the Web, Information Processing and Management, Vol. 36(2), 207-227.

27. Jansen, B. J. and Pooch, U. (2001) A Review of Web Searching Studies and a Framework for Future Research, Journal of the American Society for Information Science and Technology, Vol. 52(3), 235-246.

28. Fox, S., Karnawat, K., Mydland, M., Dumais, S. T. and White, T. (2005) Evaluating Implicit Measures to Improve the Search Experience, ACM Transactions on Information Systems, Vol. 23(2), 147-168.

29. Kelly, D. (2005) Implicit Feedback: Using Behavior to Infer Relevance, New Directions in Cognitive Information Retrieval, Vol. 19.

30. Kelly, D. and Belkin, N. J. (2001) Reading Time, Scrolling and Interaction: Exploring Implicit Sources of User Preferences for Relevance Feedback, In Proceedings of the 24th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 408-409.

31. Hilbert, D.M. and Redmiles D.F. (2000) Extracting Usability Information from User Interface Events, ACM Computing Surveys, Vol. 32(4), 384-421.

32. Firefox Continues to Erode Microsoft Dominance, http://www.netapplications.com/news.asp

33. Browser Market Share White Paper, http://www.e-janco.com/browser.htm

34. Manber, U., Patel, A. and Robison, J. (2000) Experience with Personalization of Yahoo!, Communications of the ACM, ACM Vol. 43(8), 35-39.

35. Personalization is not Technology: Using Web Personalization to Promote your Business Goal, http://www.boxesandarrows.com/view/personalization_is_not_technology_using_web_personalization_to_promote_your_business_goal.

36. Eirinaki, M. and Vazirgiannis, M. (2003) Web Mining for Web Personalization, ACM Transactions on Internet Technology, Vol. 3(1), 1-27.

37. Mobasher, B., Cooley, R. and Srivastava, J. (2000) Automatic Personalization Based on Web Usage Mining, Communications of the ACM, Vol. 43(8), 142-151.

38. Sugiyama, K., Hatano, K. and Yoshikawa, M. (2004) Adaptive Web Search Based on User Profile Constructed Without Any Effort From Users, In Proceedings of the 13th International Conference on World Wide Web, 675-684.

39. Sugiyama, K., Hatano, K., Yoshikawa, M. and Uemura, S. (2004) User-Oriented Adaptive Web Information Retrieval based on Implicit Observations, In Proceedings of the 6th Asia Pacific Web Conference, 636-634.

40. Yan, T.W. and Garcia-Molina, H. (1995) SIFT - A Tool for Wide-Area Information Dissemination, In Proceedings of the USENIX Technical Conference, 177-186.

41. My Yahoo, http://my.yahoo.com/

42. My Netscape, http://my.netscape.com/

43. Google Personalized Search, Google Labs, http://labs.google.com/

44. Shahabi, C. and Chen, Y. (2003) An Adaptive Recommendation System without Explicit Acquisition of User Relevance Feedback, Distributed and Parallel Databases Vol. 14(2), 173-192.

45. Collewijn, H. (1999) Eye Movement Recording, In R. H. S. Carpenter & J. G. Robson (Eds.), Vision research: A practical guide to laboratory methods, Oxford: Oxford University Press, 245-287.

46. Pazzani, M. and Billsus, D. (1997) Learning and Revising User Profiles: The Identification of Interesting Web Sites. Machine Learning, Vol. 27(3), 313-331.

47. Kosala, R. and Blockeel, H. (2000) Web Mining Research: A Survey, SIGKDD ACM SIGKDD Explorations Newsletter, Vol. 2(1), 1-15.

48. Sun, J., Zeng, H., Liu, H., Lu, Y., and Chen, Z. (2005) CubeSVD: A Novel Approach to Personalized Web Search, In Proceedings of the 14th International Conference on World Wide Web, 382-390.

49. White, R. W. and Drucker, S. M. (2007) Investigating behavioral variability in web search, In Proceedings of the 16th International Conference on World Wide Web, WWW '07. ACM Press, New York, NY, 21-30.

50. Hu, J., Zeng, H., Li, H., Niu, C., and Chen, Z. (2007) Demographic prediction based on user's browsing behavior, In Proceedings of the 16th International Conference on World Wide Web, WWW '07, ACM Press, New York, NY, 151-160.