Analysis of Web Usage Patterns in Consideration of Various Contextual Factors

Jinhyuk Choi
Korea Advanced Institute of Science and Technology (KAIST)
119, Munjiro, Yuseong-gu
Daejeon, 305-732, Republic of Korea
demon@kaist.ac.kr

Jeongseok Seo
Information and Communications University (ICU)
119, Munjiro, Yuseong-gu
Daejeon, 305-732, Republic of Korea
chaoticblue1@icu.ac.kr

Geehyuk Lee
Korea Advanced Institute of Science and Technology (KAIST)
119, Munjiro, Yuseong-gu
Daejeon, 305-732, Republic of Korea
geehyuk@kaist.ac.kr

Abstract

It is important to analyze user's Web usage logs for developing personalized Web services. However, there are several inherent difficulties in analyzing usage logs because the kinds of available logs are very limited and the logs show uncertain patterns due to the influences of various contextual factors. Therefore, speculating that it is necessary to find what contextual factors exert influences on the usage logs prior to designing personalized services, we conducted several experiments in-series not only in situations of performing designed tasks during short time periods but also in users' natural Web environments during a period of several days. From the results of our experiments, we found that interest levels, credibility levels, page types, task types, and languages are influential contextual factors in a natural Web environment. Moreover, some historical and experiential patterns that could not be observed in short time analysis were discovered in the results of long time analysis. These findings will be useful for other researchers, practitioners, and especially for developers of adaptive personalization services.

Introduction

The World Wide Web has a unique characteristic in that the amount of contained information is continuously increasing and yet can still be reached easily by users through various Web services. Moreover, it provides various types of media so that users can use it for multiple purposes. Therefore, it is very important for researchers and practitioners to make the Web even more effective for finding necessary information.

One of various means by which we can make the Web more useful is to develop intelligent information delivery in order to allow users to find their target information more effectively. A core part of intelligent information delivery is to search through personalized contents without the user's explicit participation. For personalization, it is necessary to learn more about the user and to build a user model based on this knowledge. This personalization process is a main topic of research on Web usage mining (Mobasher et al. 2000; Gauch et al. 2007). However, it is not easy to learn more user information because we cannot explicitly ask the user about his/her characteristics or what he/she is thinking at any particular time we want to know. This means that we have to find another way to learn more information about them. From this perspective, many researchers have looked for effective implicit methods to learn more about users, and many intelligent methods have been actively suggested by several researchers (Kelly and Teevan 2003; Kelly 2004; Kelly and Belkin 2004; Kelly and Cool 2002; Choi et al. 2007; Hofgesang 2006; Seo and Zhang 2000; Badi et al. 2006; Al halabi et al. 2007; Kellar et al. 2005). In their researches, usage logs that are stored while users visit Web pages have been used to learn about particular user interests. For examples, the URLs of visited Web pages, visit period, dwelling time, mouse clicks, mouse movement, keyboard typing, and visit frequencies on each Web page have been applied as implicit interest indicators.

Although many successful results have been provided so far, there are several inherent difficulties in analyzing usage logs and extracting necessary information from them. The first difficulty comes from the fact that the kinds of available usage logs are very limited, and there are no standard ways to interpret the meaning of usage patterns. This means that we have to carefully investigate usage patterns prior to using the logs as effective indicators. Secondly, Web users are under the influence of various contextual factors while they use the Web, as it has multiple aspects as a simple information tool, social communication mediator, entertainment source, and so on. Therefore, usage logs will show very uncertain patterns because various contextual factors will exert their influence on the usage patterns concurrently (Kelly and Belkin 2004). The third difficulty is related with the historical aspect in that a user's experiences also exert influences on the variation of usage patterns. Therefore, a Web usage pattern analysis should be a long-term process because it cannot be adequately performed by studying only short-time usage. In addition, to analyze a user's various characteristics, the usage data should be collected at the browser side in the user's real Web environment for a long period without any constraint on a specific Web server.

This paper details the results of our experiments in which we initially tried to find the possibilities of overcoming the above difficulties. For our experiments, various usage logs have been collected at the browser side and carefully analyzed not only in situations of performing designed tasks during short time periods but also in users' natural Web environments during a period of several days. We obtained several interesting findings from the results. We think that these findings will be useful for other researchers, practitioners, and especially for developers of personalization services.

This paper is organized as follows. In section 2, we review some of the related researches. In section 3, we describe our experimental procedure and the results that have been obtained so far are given in section 4. In section 5, summary and future works are introduced.

Related Work

Human Information Behavior

There have been a lot of studies that have focused on human information behavior analyses in various research fields. In those studies, the researchers have focused on several contextual factors that affect a user's behavior, conceptualizing the relationships between information-seeking behavior and contextual factors. In (Sonnenwald 1999), the authors proposed an evolving framework in which cognitive, social, and system perspectives are incorporated. In the framework, human information behavior including information exploration, seeking, filtering, use, and communication were included. Based on the framework, various influential factors - physical, cognitive, affective, economic, social, and political – and their implications were investigated. In (Johnson 2003), the needs of an information-seeking behavior analysis in a multi-contextual environment were presented and a theoretical framework was suggested. The authors of (Kari and Savolainen 2007) asserted that users are also improving along with the change of information environment, and they found 11 relationships between individual developmental objectives and information searching via the Internet. In (Byström and Järvelin 1995; Borlund and Ingwersen 1997; Bystrom 2002; Vakkari 1999; Vakkari 2001), the influence of task complexity on information seeking behaviors was investigated. An overview of the nature of trust, and a framework of trust-inducing interface design features, were given in (Wang and Emurian 2005). Particularly in (Wang et al. 2000), the authors introduced a multidimensional model of user-web interaction, and three dimensions – user, interface, and the Web – were considered. In the model, the user dimension is considered to be influenced by the particular task, information need, knowledge state, cognitive style, affective state, and so on. They measured users' cognitive styles and affective states before a user study, applying a process-tracing technique while users were conducting information-seeking tasks, and found various types of relationships among the elements of the dimensions. In (Fogg et al. 2003), based on results of an online qualitative study, the credibility for Web contents were considered and important factors of credibility were suggested to formulate Web design guidance. In (Wathen and Burkell 2002), the authors asserted that users filter out most of the gathered information and retain only useful information. In addition, they concluded that the credibility or believability of information is one of the most important criteria for the filtering. In (Rieh 2002), the authors found that users judge cognitive authority and information quality by two types of judgment - predictive judgment and evaluative judgment – and they also identified the main facets and keywords of the judgments through a user study.

From these researches, we found that human information behavior cannot be studied without the consideration of the influences of various types of contextual factors. However, because the purposes of these researches were not to develop an intelligent system but to construct theoretical models, they did not study quantitatively how the Web usage patterns reflect the influences of various contextual factors.

Web Usage Mining

There has been a lot of effort to quantitatively measure the influences of contextual factors on Web user behaviors based on various usage logs in the field of Web usage mining. Among the various factors, user interest toward content has been the main focus of researchers. The various implicit indicators of user interest can be found in (Kelly and Teevan 2003). In (Kelly 2004), the familiarity of a topic has been discussed, and the authors concluded that as one's familiarity with a topic increases, his/her searching efficacy increases and reading time decreases. For user characteristics, cognitive and problem-solving styles were studied in (Kim and Allen 2002). In their study, the authors observed various user activities - average time spent, average number of Websites viewed, average number of bookmarks made, and average number of times a search/navigational tool was used for completing a search task – while the users performed two types of given tasks in an experimental environment, and the authors found that there are significant differences among user activities according to the type of task and user's problem solving style. For usage logs, the display time was discussed most actively. In (Kelly 2004; Kelly and Belkin 2004), based on gathered data from 7 subjects for 14 weeks, the relationships between display time and various factors – task, topic, usefulness, endurance, frequency, stage, persistence, familiarity, and retention – were investigated, and the authors concluded that the display time is not suitable for inferring a user's interest because there is great variation between display time and interest according to the user; large differences according to the task at hand also appear. On the contrary, in (Choi et al. 2007), the viewing time has been used as a good implicit indicator, and in (Hofgesang 2006), the authors made an assertion that time spent on a Web page is more important than visit frequency in inferring a user's interest. In (Seo and Zhang 2000), bookmarking, time for reading, following up the HTML document, and scrolling were used as relevant activities, and a machine learning algorithm was applied to learn the user's characteristics. In (Badi et al. 2006), various parameters of document attributes, document reading activities, and document organizing activities were investigated to recognize user interest and document values. In (Kellar et al. 2005), the authors found that the time spent is more useful for more complex Web searching tasks. In (Nakamichi et al. 2006), the authors also used several quantitative data of user behavior – browsing time and moving distance, moving speed, and wheel rolling of the mouse – to detect low usable Web pages.

Most of the researches have analyzed usage logs with the intention of developing an intelligent system that learns user characteristics and builds a user model. However, most of the studies did not fully consider the influences of various contextual factors, or they focused only on a user's interest without consideration of other types of subjective feedback together. Moreover, most researches except (Kelly, 2004; Kelly and Belkin 2004) did not consider the historical aspects of usage data that can only be gathered by a long-time analysis in a user's natural Web environment.

Our Approach

Before everything else, we reviewed previous related researches carefully and collected contextual factors for consideration and usage logs that can be obtained at the browser side. The contextual factors and usage logs that we considered are given in Table 1.

We carried out not only a qualitative analysis but also a quantitative one. For ecological validity, we also observed users in their own personal places. Because some of the contextual factors are inherently subjective and cannot be measured with only usage logs, we collected various types of feedback regarding the current context directly from users. However, to minimize the burden on the users in this study, we tried to minimize the number of feedback questions as much as possible. We developed software that runs on each user's PC in order to collect their behavior logs and feedback in their Web browsing environments.

Contextual Factor

Contextual factors include subjective assessments about contents, situational factors, a user's individual characteristics, and so on. Because these factors cannot be measured systemically, we designed a process in which we can obtain the users' subjective feedback directly. First, we

	Task	**Contextual factors**	**Usage logs**	**Period**
Ex1	Visit collected pages (text only)	Interest	Viewing time Mouse movement Mouse wheel Mouse clicks WM_PAINT	2 hrs
Ex2	Visit collected pages	Interest Complexity Difficulty Credibility	Viewing time Mouse movement Mouse wheel Mouse clicks WM_PAINT	2 hrs
Ex3	Free visits / given tasks	Interest Complexity Difficulty Credibility Task type	Viewing time Mouse movement Mouse wheel Mouse clicks WM_PAINT	2 hrs
Ex4	Free visit / free tasks	Interest Credibility Task type	Viewing time Mouse movement Mouse wheel Mouse clicks Keyboard typing Visit frequency Day frequency	2 wks

Table 1. The environment and gathered data of experiments

considered the users' attitudes toward the current task as one of the contextual factors. Actually, the types of user task can be classified into detailed categories – information seeking, fact-finding, transaction, and browsing (Kellar et al. 2007). However, we classified user tasks into only two categories – careful searching and casual searching - according to the users' attitudes toward the current task. A detailed description of the task categorization appears in section 3.5. There are more contextual factors that cause users to interact with Web pages. For example, a user may stay for a relatively long time at a specific Web page because there are interesting contents there, or the user feels that the contents are more useful than others. Sometimes, the user may roll the mouse wheel more frequently on one Web page than on others because he/she wants to read the entire content of the page carefully. In this regard, we selected some further factors that may exert an influence on user interactions with Web pages. The factors are interest, credibility, complexity, and difficulty. The complexity factor tells us how users feel about the layout structure of a Web page, and hence it may include a user's subjective viewpoint of usability and familiarity. We also included the difficulty factor because we thought that user behavior is subject to variation according to a subjective assessment of the difficulty of the contents displayed.

Web Usage Log

Implicit user interest analysis has shown good performance at the server-side especially for commercial Websites. However, in spite of the fact that it is easier to analyze user interest at the server-side, currently many researchers have focused on browser-side analyses because user interest can be analyzed from various Websites, and a user model can

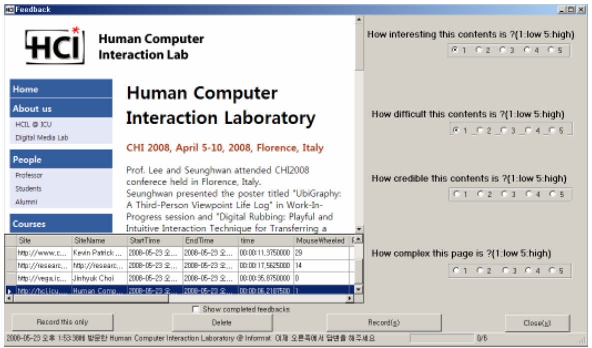


Figure 1. The feedback window consists of a browser control to view the contents of visited web pages, a list window to choose a visited URL, radio buttons to choose the answer of some questions, and so on

be constructed using a wealth of information through a browser side analysis. In order to analyze users' implicit interest at the browser side, we have to monitor several usage logs, for example, the viewing time, scroll movement, sequences of visited URLs, keyboard typing, and so on. In our research, we have chosen several usage logs to record while users view different Web pages. The viewing time that has mainly been investigated in the related researches so far is the time during which users remain on a particular web page. The mouse wheel counts the number of WM-MOUSEWHEEL messages (Choi et al. 2007). For mouse and scrollbar movement, we measured the distance between two consecutive positions of the mouse cursor and scroll bar at regular intervals and summed the distances. We also counted the number of processed WM-PAINT messages, as WM-PAINT messages are processed when users change the size of their browser window, scroll within the window, move their mouse cursor, and so on. The number of mouse clicks and keyboard typing were also considered. We believe that these activities are good indicators of user interest regarding the contents of Web pages. We have chosen these logs because they can be measured without much effort. However, for scroll movement, we were unable to obtain the position of the scrollbar on some of the Web pages, and the WM-PAINT messages can be affected by the dynamic content of certain Web pages. This means that we have to be careful when using these data as logs for measuring user activities.

We did not record some of the behaviors that have been considered by other researches – bookmarking, saving, printing, and coping and pasting – because users do not always show those behaviors on every valuable Web page, and hence their records do not suit our purpose.

We collected some physical data of Web pages - the scroll height, file size, and URL information (top-level URL and depth of URL) - of each visited Web page. Moreover, in the course of the experiments, some additional factors were included when they were required for analysis. The additional factors were the number of out-links on a Web page, the Web page type, and the language presented (e.g., Korean or English).

We also considered carefully some historical factors that can be analyzed only through relatively long periods of monitoring. The historical factors include visit frequencies and day frequency. Among those factors, day frequency is a new concept that has not been introduced before. A detailed description of day frequency will be given in a later section.

Data Collection Software

In some of the previous researches, custom-built browsers have been used (Kellar et al. 2007), as have some specialized logging software that works "in stealth mode" (Kelly and Belkin, 2004). Although there are several merits in using custom-built browsers, because various data can be collected easily, we developed a browser-monitoring module (BMM) that runs behind Internet Explorer without any modification to the browser, as we wanted to preserve the natural state of the Web browsing environment as much as possible.

BMM is a type of monitoring software that was developed to detect Windows GUI messages while users read Web pages, and thus it is possible to measure user activities in real-time without any interruption to the users. BMM uses a global hooker library, written in C++, which runs in the background and hooks all Windows operating system events. In addition, using Windows Shell API, BMM can access all instances of currently running Internet Explorers through the COM object. In addition, necessary properties of Web pages can be obtained from the COM object. BMM is written in C#, running under a Windows platform with .NET Framework 2.0.

BMM consists of four components - hooker, data recorder, data aggregator, and feedback window. The data to hook are the number of keys pressed, events of program focus changes, number of WM_PAINT events, mouse click and mouse wheel messages, and so on. Basically, the hooker catches every message passed within the operating system, so we should filter out irrelevant messages to record only necessary data for our studies. For instance, because a WM_PAINT message is invoked whenever the O/S needs to re-draw some parts of a window, we have to be able to ignore the messages from unfocused windows and count the number of messages that are invoked for only the currently focused browser window. The aggregator can acquire several properties of a Web page by using a Document Object Model (DOM). Acquired properties are the viewing size of a document (in pixels), file size (in bytes), current location of the scrollbar, and character set of the page. The location of scroll bar is periodically updated so that the total displacement of the scrollbar can be estimated. However, a critical issue arises at several 'fancy' Web pages that have different structures from standard Web documents, eventually yielding no data while accessing the DOM property. The data aggregator also aggregates all data from these multiple components, and the data recorder stores the aggregated data in a human-readable XML format for future analysis. After Web searching, using the feedback window, users can review the visited Web pages and choose radio buttons that ask about several types of assessments about the contents of each Web page. If the users do not want to answer questions regarding some of the Web pages, they can even remove the records easily. In figure 1, the structures of the feedback windows are shown.

Subject

We conducted 4 experiments, each with its own purpose. The detailed concept of the experiments will be described in the next section. For each experiment, we recruited some graduate students who are majoring in computer science for our subjects. Twenty-five students participated in the first experiment, 23 in the second, 19 in the third, and 12 students in the fourth. Among the students, 11 got through the second, third, and fourth experiments, and one new subject volunteered for the fourth experiment. All of the students have a high level of knowledge and experience about the Internet and the Web. We chose these students as subjects because all of them use the Web not only for their work but also for entertainment or distraction. Most of all, they use the Web for a relatively long time each day so that we could gather plenty of data from their activities. It also means that we could observe their Web usage patterns under various contexts. We paid about 20 dollars to each subject for their participation in the first, second, and third experiments, respectively. For the fourth experiment, we paid 60 to 160 dollars to each subject according to the rate of the completed feedback.

Experimental Concept and Procedure

There are three main strategies for studying information-seeking behavior – laboratory experiments, sample surveys, and field studies (Kellar et al. 2007). Considering these strategies, we designed four experiments and conducted them in-series. In the first and second experiments, the subjects came to our laboratory and browsed some pre-collected Web pages. In the third experiment, the subjects performed given information-seeking tasks in our laboratory. As a final step of each experiment, the subjects carried out feedback tasks in order to record their own subjective assessments about each of the Web pages they had browsed. The fourth experiment was carried out at the subjects' own residences. The subjects installed BMM on their PCs to collect their Web usage logs for a period of about two weeks. For the feedback process of the fourth experiment, we let the subjects carry out the feedback tasks at least once a day. The first and second experiments were carried out in a blind mode in which the subjects could not see any information about the contents of each Web page before viewing them. In other words, no proximal cues (Chi et al. 2001) were provided.

Experiment 1. The first experiment was a kind of preliminary study. We collected 120 Web pages that contain only text and offer information on various topics – politics, economics, education, engineering, entertainment, science, health, and sports – with varying content size. The twenty-five subjects read each page in their own desired manner from the list of collected Web pages. Because we wanted to exclude any effect of information clues, we simply provided numbers on the list without showing any information about the contents of the Web pages in advance. Thus, the subjects were supposed to click the numbers in order to view the contents. To obtain the appropriate data, the subjects were not told that some activities would be measured while they were viewing the Web pages. During the experiments, the subjects' activities while reading the Web pages, and some measurable data, were recorded in a log file for future analysis. In addition, whenever a subject finished reading a Web page, a small window appeared wherein the subject recorded his/her interest level for the contents of the page. There were 5 levels of interest, and the subjects recorded their interest for the contents of a Web page accordingly. Due to some malfunctions of the BMM in the users' browsing environment and a failure to properly obtain user feedback, the log files of 5 users were excluded. Therefore, we analyzed 20 users' log files. For the first experiment, we formulated the following simple hypotheses.

1. The number of processed log data is relatively higher on Web pages that contain interesting contents.

2. The amount of information in a Web page affects the amount of processed log data.

Experiment 2. Actually, the procedure of the second experiment was the same as the first experiment except that we collected ordinary Web pages that contain images, tables, videos, and frames. It was intended to see whether there will be differences in usage patterns according to form of the Web page. When a subject finished reading all of the Web pages, he/she activated a feedback window wherein the subject could review all of the pages and answer some questions about each one visited. In this experiment, differently from the first experiment that collected only the interest levels for the contents, we also wanted to verify the influence of other subjective assessments of Web pages - difficulty, complexity, and credibility along with interest – on a 5-point scale. If a subject clicked one of the URLs on a visited page list in the feedback window, the contents of the Web page appeared again, and the subject could then choose his/her points for the questions regarding the subjective feedback.

Experiment 3. We can find several different categorizations of Web user behaviors in previous researches. Most recently, 4 task categories were provided in (Kellar et al. 2007) - fact finding, information gathering, just browsing, and transactions. In (White and Drucker 2007), Web users are grouped into navigators and explorers according to the level of visit variances. In consideration of these previous works, we also classify a user's Web tasks into two groups.

Ex	Feedback	VT	MM	MW	MC	WP
Ex1	interest	0.695 (**)	0.572 (**)	0.563 (**)	0.475 (**)	0.663 (**)
	scroll height	-0.006	0.006	0.261	0.008	0.059
Ex2	interest	0.771 (**)	0.545 (**)	0.686 (**)	0.559 (**)	0.507 (*)
	complexity	-0.391 (**)	-0.178	-0.148 (**)	-0.196	-0.599 (*)
	difficulty	-0.476	-0.340	-0.532	-0.418	-0.057 (**)
	credibility	0.507	0.203	0.411 (*)	0.289	0.241
	scroll height	0.074	0.016	0.167	0.001	-0.059
Ex3	interest	0.396 (*)	0.301 (*)	0.119	0.245	0.229
	complexity	-0.315 (**)	-0.129 (**)	-0.533 (**)	-0.162 (**)	0.040
	difficulty	0.307	0.307	-0.124	0.182	-0.330
	credibility	0.609 (**)	0.414 (**)	0.288 (*)	0.412 (**)	0.389
	scroll height	0.011	-0.025	0.120	-0.036	-0.022
Ex4	interest	0.442 (**)	0.315	0.258	0.306	0.282 (KT)
	credibility	0.434 (*)	0.138	-0.010	0.222	0.124 (KT)
	scroll height	0.056	0.001	0.117	0.017	0.001 (KT)

VT: Viewing time / MM: Mouse move / MW: Mouse wheel / MC: Mouse click / WP : WM_PAINT / KT: Keyboard typing
*: p-value of ANOVA test < 0.05
**: p-value of ANOVA test < 0.01

Table 2. The values of correlation between feedback level and the amount of usage logs

Task 1: careful searching

This task is a type of information gathering that requires accuracy, trust, efficiency, and responsibility of the search results. In our experiment, the given task was to find some information about their research topics. For examples, they had to find some Web pages of laboratories in universities or companies that are related with their research topics and read the pages carefully to judge the relevance of the information. We encouraged the subjects to perform this task as normally as possible.

Task 2: casual searching

This task is a type of information gathering and browsing that can be performed without any burden or responsibility regarding the search results. For example, the subjects could search for some information about their hobbies, favorite products to buy, famous tourist spots, favorite sports or movie stars, and so on. We also encouraged the subjects to perform these tasks as normally as possible.

The subjects performed the two tasks with their own topics for about 2 hours. The logging data and feedback

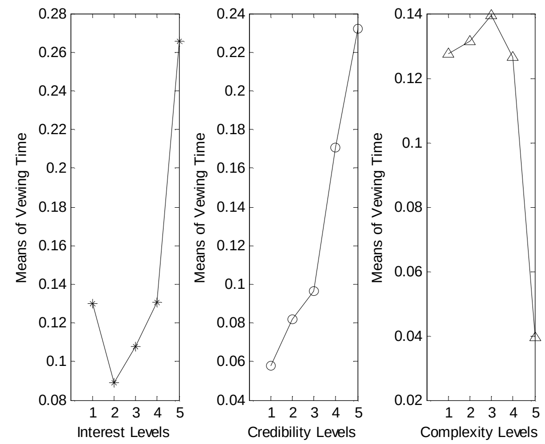


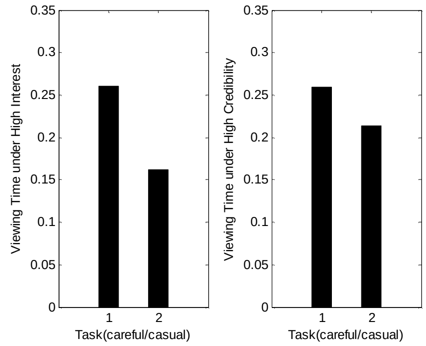
Figure 2. The viewing time according to feedback levels in the third experiment

details were the same as in the second experiment. Differently with the first and second experiments that controlled the subjects' activities in that the subjects could only visit the collected Web pages without any pre-information clues, in the third experiment, the subjects could visit any Web page that they wanted and use any search engine or portal site they wanted to use. Therefore, we observed a lot of re-visitation patterns. Thus, during the feedback phase, we let the subjects delete the logs of Web pages that they just used to find other Web pages to visit. In this way, we excluded the navigational Web pages. The concepts of the navigational Web pages will be given in section 4.5.

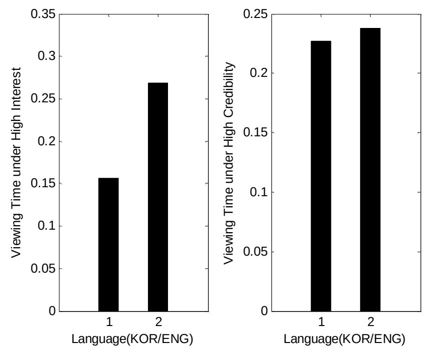
Experiment 4. For the fourth experiment, 12 graduate students participated - 4 females and 8 males. They installed the BMM on their PCs and collected various logs for about 2 weeks. Some of the subjects participated in our experiment for 16 days. For their feedback, we encouraged them to give their feedback levels of each visited Web page a 5-point scale and choose one of the task types. If a URL was not a content page according to the subject's viewpoint, the URL could be deleted easily and BMM records a special number for the URL for future analysis. In this experiment, we collected only three types of feedback – interest, credibility, and task types - because we wanted to minimize the subjects' burden in answering many questions for all of the visited Web pages visited.

Result

In the series of experiments, we measured the numbers of several processed messages on each visited Web page and normalized the value using min-max normalization according to each subject. We included this normalization procedure because there would be variances in the amount of usage logs due to the subjects' individual differences.



(a)



(b)

Figure 3. (a) The differences of viewing time according to task types (b) the differences of viewing time according to languages

Experiment 1

From the results of the first experiment, we found some interesting patterns. As we can see in table 2, there were positive correlations between the amount of all usage logs and interest levels. Furthermore, from a one-way ANOVA test, we also found that the amount of the logs shows significant differences among the interest levels. Based on this result, we temporally concluded that users have a tendency to interact more at high-interested Web pages, and hence all the logs can be used as implicit interest indicators. One more interest thing is that there was a low correlation between the amount of usage logs and the size of the Web pages except for the amount of the mouse wheel log.

Experiment 2

Actually, we thought that there would be some differences between the result patterns of the first experiment and those of the second experiment because the forms of the Web pages were quite different. However, there were no big differences between the results. Table 2 shows us that there were also positive correlations between the amount of all usage logs and interest levels, similarly with the results of the first experiment. In addition, we also found significant differences in the amount of usage logs among the interest levels. This means that the form of the Web pages is not an important factor. Differently from the results of interest level, the difficulty and complexity levels showed a negative correlation with the amount of usage logs. The credibility levels showed positive correlations with the amount of usage logs but the differences of the amounts among the levels are not statistically significant. From the results, we concluded that the interest level exerts the most significant influence on the amount of usage logs, and that users are inclined to quickly leave Web pages that have difficult contents or complex structures without many interactions. Finally, we found that there were low correlations between the amount of usage logs and the sizes of Web pages except for the amount of the mouse wheel log. This was not different with the results of the first experiment.

Experiment 3

In figure 2 and table 2, we can see that the viewing time and amount of mouse movement have positive correlations with the interest levels, and that the differences of the amounts among the interest levels are also statistically significant. The amount of mouse wheel use, mouse clicks, and processed WM_PAINT messages also showed positive correlations with interest levels, but the differences were not statistically significant. The amount of usage logs increased according to the complexity levels, but dropped steeply at level 5. The difficulty levels showed no large correlation with the amount of usage logs. The most interesting pattern that we found in the results of the third experiment was that the amount of usage logs showed a positive correlation with the credibility levels, and that the differences of the amounts of usage logs among the credibility levels were statistically significant. This result was not found in the results of the second experiment in which users browsed pre-collected Web pages without proximal cues. Therefore, we concluded that the usage logs are under the influence of credibility levels as well as interest levels in ordinary Web browsing environments.

In the third experiment, we also checked whether there are differences in the amount of usage logs according to the task types and written languages used. From figure 3, we found that there was a general trend of more interaction logs recorded during a careful task than during a casual one, especially on pages of the highest interest and credibility levels. For written languages, there was a general trend of more interaction logs on English pages than on Korean pages, especially on the pages with the highest interest levels, but there was no large difference according to credibility levels. These results showed us that the type of task and written languages used also should be considered as important influential factors that make differences in the amount of usage logs created.

Experiment 4

In the fourth experiment, there were some logs that contain an excessively long viewing time because the experiment has been conducted in the users' personal environments.

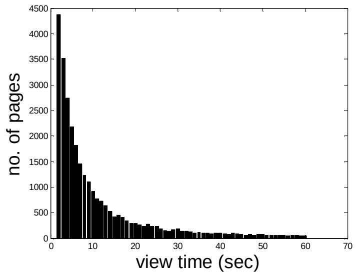


Figure 4. The distribution of viewing time

From figure 4, we can find that the users stayed on 99% of the all visited Web pages for at most 346 seconds. We also found that there were some visited logs that showed a viewing time of over 30 minutes. This means that we should find a maximum cutline in order to filter out some logs as simply noise. We set various values to the cutline, from 3 to 25 minutes. As we set the cutline values differently, we excluded logs in which the viewing time was above the cutline, and then normalized each user's viewing time to his/her scale. Finally, we checked whether the magnitude of the cutline made an impact on the applicability of the viewing time as an indicator. From figure 5, we can see that a reasonable cutline should be set to somewhere between 5 and 18 minutes in order to observe a high positive correlation between the viewing time and interest level, and the statistical difference among the viewing times in each interest level. For example, if we set the maximum cutline to 14 minutes, the viewing time shows a positive correlation with the interest level ($r = 0.5522$), and according to the result of a one-way ANOVA test, the differences among the viewing times of each level are significantly different ($p = 0.0092$). This means that we can use viewing time to identify interested Web pages based on the fact that users will stay for a relatively longer time on them than on uninterested Web pages. In addition, we found that when we want to infer users' interest based on the viewing time, a careful noise-filtering task is absolutely required. Therefore, we excluded logs that contained over 15 minutes of viewing time in the fourth experiment. In figure 6 and table 2, we can see that only the viewing time showed positive correlations and statistically significant differences among the levels of interest and credibility. It is very interesting that we could not find significant differences between other usage logs and feedback levels. The differences in the amount of usage logs according to the task types were similar with the result of the third experiment.

Additional Findings from Experiment 4

Because the fourth experiment was conducted during a period of about 2 weeks, we can observe some more historical patterns that could not be observed in previous experiments. In this section, we introduce some additional findings.

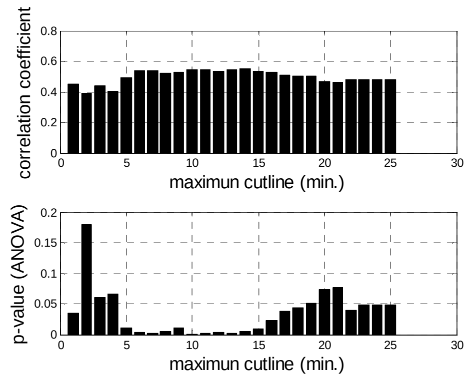


Figure 5. The correlation coefficients (top) and p-values of significance test (bottom) according to maximum cutline

Day frequency and feedback pages vs. non-feedback pages. Because most of target pages that users want to access can be reached via portal sites, news sites, and search engines, we thought that the front pages of these sites and hub pages within the sites may appear in the visited URL history more frequently than others. For example, when a user wants to read a newspaper, he/she visits the home page of news site and clicks on some links that seem to contain interesting news. In a similar manner, whenever a user wants to find some information, he/she may visit the front page of a search engine first and then click on one of the links that the search engine retrieves. Similarly, if the user wants to log onto some commercial sites or even his/her own Web mail accounts, he/she should first visit the front page of the service and input his/her username and password in order to proceed. Therefore, if we look over the users' visited URL histories, the navigational pages - the front pages of portal sites, news sites and search engines, and any type of hub page – will appear more frequently than others. Moreover, if the users visit Websites according to their daily routine, they will visit some of the Websites everyday in their regular patterns. In this respect, we thought that the URLs of navigational pages might be found in logs from each day. On the contrary, the content pages were shown relatively

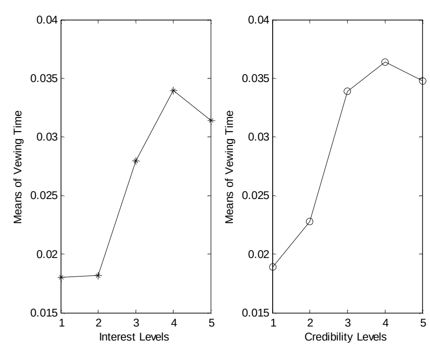


Figure 6. The viewing time according to feedback levels in the fourth experiment

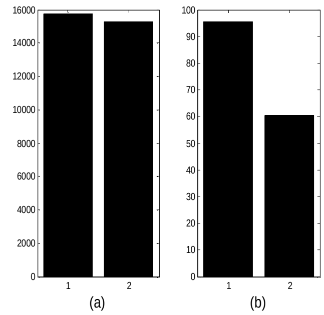


Figure 7. (a) The number of feedback pages and non-feedback pages and (b) the average number of outlinks contains: 1 – feedback page / 2- non-feedback page

rarely because the users don't usually view the same contents again and again.

Based on the considerations that we have mentioned so far, we formulated a very simple hypothesis - everyday-visited URLs have a strong chance to be navigational pages. For the hypothesis, we created a variable named Day Frequency (DF). The concept of DF is very similar to document frequency, which is often used in information retrieval and text mining (Salton and McGill 1986), and DF value of each visited URL can be calculated using equation (1).

$$DF_i = \frac{|\{d_j : Url_i \in d_j\}|}{|D|} \quad (1)$$

In this equation, $|D|$ is total number of days in experiment, d_j is the URL collection of the j-th day and $|\{d_j : Url_i \in d_j\}|$ means the number of days where i-th URL appears. If a URL exhibits a high value of DF, the URL is thought to be inappropriate for content extraction and should be regarded as a navigational page.

In the fourth experiment, the selection of a contents page was fully up to the subject's subjective decision. Even though we did not explain the concept of navigational pages in detail, they found by themselves that there are naturally several Web pages that may not be fit for expressing their feedback levels. As we can see in figure 7, the number of non-feedback pages was much greater than our expectation. The 12 subjects have mainly deleted the home pages of search engines, retrieved lists of search engines, the first pages of portal sites, news lists, home pages of community sites, online banking sites, intranet front pages and so on as non-feedback pages. In some of the previous researches, we found that there were several attempts to discriminate content pages from navigational pages using the number of outlinks that are contained in the pages (Cooley et al. 1999; Fu et al. 2001; Domenech and Lorenzo 2007). The main idea is that there will be a larger number of outlinks on navigational pages than on contents pages. We also thought that this idea is acceptable so we counted the average numbers of contained outlinks in both feedback pages and non-feedback pages. However, as we can see in figure 7, the number of outlinks on feedback pages was higher than on non-feedback pages in our results. Therefore, we examined carefully whether the DF values in feedback pages and non-feedback pages are significantly different. As we can see in figure 8, the average DF value of non-feedback pages is higher than the values of feedback pages, and the difference is statistically significant (p = 0.0003). We found that the amount of some usage logs was also different between feedback and non-feedback pages. From table 3, we can see that viewing time, the amount of mouse wheel use, and the amount of keyboard typing were significantly different.

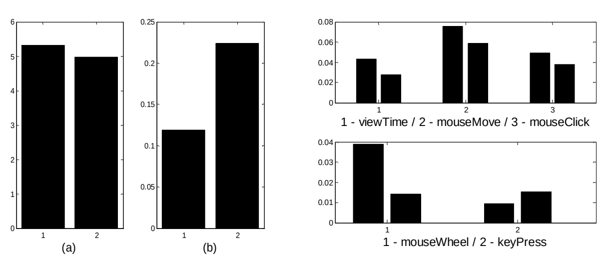


Figure 8. The URL depth of feedback pages and non-feedback pages (left - a) and the DF values (left - b) : 1 – feedback page / 2- non-feedback page and the mean values of interaction logs: on feedback pages (right - left bars) and on non-feedback pages (right - right bars)

logs	p-value
URL Depth	0.0623
Day Frequency	0.0003 (*)
Viewing Time	0.0206 (*)
Mouse Move	0.5314
Mouse Click	0.5258
Mouse Wheel	0.0181 (*)
Keyboard typing	0.0349 (*)

Table 3. The results of significance test for difference of the values of each interaction log between feedback pages and non-feedback pages: (*) means significant

Task Identification by Visited URLs. We believed that users have their own URL lists that are specific to their current tasks because they may use the Web based on their individual previous experiences on the Web. In this respect, we analyzed the top-level URLs that users visited during the period of the experiment. As we can see in table 4, over 90% of visited URLs were separable by the tasks.

user No.	task separable (%)
1	92.68
2	92.59
3	93.17
4	95.77
5	75
6	93.86
7	100
8	90.57
9	89.29
10	97.40
11	91.07
12	96.21

Table 4. The proportion of task separable URLs

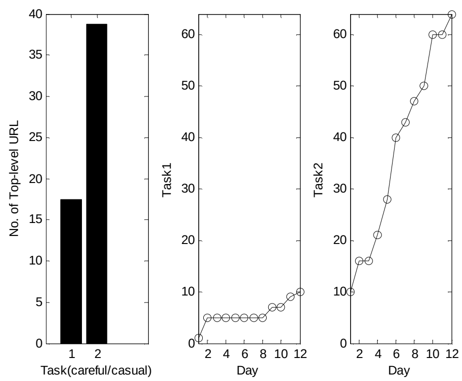


Figure 9. The average number of URLs in each task (left), the increasing rate of average number of URLs in careful task (middle), in casual task (right)

In other words, 90% of visited URLs belong to a specific task only, and hence we can infer the types of current task easily by checking the top-level URLs. Moreover, as we can see in figure 9, the number of URLs that users visited in the casual tasks is much higher than in the careful tasks. The most interesting patterns are the increasing rates of the number of visited URLs as time goes on. The number of visited URLs in the tasks of casual searching increased more drastically than in careful searching. This means that the subjects showed the navigator's patterns in careful searching tasks but showed the explorer's patterns in casual searching tasks (White and Drucker 2007). We believe that this pattern is meaningful in developing personalization schemes that are adaptive to current task types.

Discussion and Future Work

Review of the Result and Summary

We analyzed the results of 4 experiments and recognized that there are noticeable differences in usage patterns according to the experimental environment. In this section, we briefly summarize the interesting differences.

Experiment 1 vs. Experiment 2. The forms of the Web pages that the subjects visited in the first and second experiments were different, but we could not see large differences between the results of the two experiments. Moreover, the amount of usage logs was not influenced by the amount of contents or size of the Web pages. We believe that this pattern came from the fact that Web users read Web pages in a nonlinear pattern, and that there are some unique characteristics in reading digital documents (Liu 2005).

Experiment 2 vs. Experiment 3. In the results of the third experiment in which the subjects freely select the Web pages to visit, we observed that the credibility levels regarding the contents exert a noticeable influence on the amount of usage logs, but the same pattern has not been observed in the results of the second experiment in which the subjects visited pre-collected Web pages even without any pre-clue about the contents. In addition, there were significant differences among the amounts of all usage logs according to interest levels in the results of the second experiment, but only the amount of viewing time and mouse movements were affected by the interest levels in the results of the third experiment.

Experiment 3 vs. Experiment 4. Differently from the results of the third experiment, we observed that the viewing time only showed a significant relation with the interest and credibility levels in the results of the fourth experiment. This means that the more natural the environment is, the more unknown factors will exert their influences on the usage patterns. We also observed in the results of the third experiment that there are some differences in usage patterns according to the task types, such that the amount of usage logs on interested Web pages in careful tasks is higher than in casual tasks. The same result was observed in the fourth experiment. Finally, from historical data analyses, we found that Day Frequency and some usage logs are significantly different according to the page types.

Summary. We also briefly summarized all of the observed patterns as the following.

1) Generally, the amount of usage logs is not under the influence of the size and form of the Web page.

2) Information scents exert noticeable influence on usage patterns such that Web users choose links to visit based on information scents, and the scents also cause the users to show some uncertain usage patterns while they are viewing Web pages.

3) The viewing time is the best log to be used as an implicit feedback indicator if it is pre-processed carefully. It means that we have to analyze the viewing time more carefully than other logs to develop personalization services that are adaptive to user interest.

4) The viewing time is under the influence of interest and credibility levels. In other words, interest and credibility levels are the most influential contextual factors in a natural Web environment. The difficulty and complexity levels do not create noticeable variations on the amount of usage logs.

5) The viewing time is also under the influences of current tasks, written languages, and page types. In addition, page types are also influential on the variations of other usage logs such that the amount of mouse wheel use, number of visits in a day, and the amount of keyboard typing were significantly different based on the page types.

6) Web users visit different Websites when they are performing different tasks and they show different navigational patterns according to the task types.

7) We recognized that some historical and experiential aspects that may not be observed in short time analysis can only be found in long time analysis.

Limitations of the experiments

Although many interesting patterns were observed from our experiments, we also acknowledge the limitations of our experiments. We cannot expect that the observed patterns will generalize to a general population because we recruited small number of people from same population for our subjects according to our experimental convenience. However, the results show us valuable usage patterns of experienced Web users and consequently provide us with a good insight into further researches.

Future Work

As we already discussed in previous sections, the viewing time is under the influence of various factors. We cannot decide what service applications are to be activated based solely on the fact that viewing time increases on a current Web page, because the viewing time will be affected by various factors - interest levels, credibility levels, page types, tasks, and written languages. Therefore, to find a user's characteristics and select the applications accordingly, it is necessary to intelligently detect what factors are currently influencing the usage patterns. We think that it will be very challenging to find current contextual factors intelligently, but we also think that the current factors can be identified through some careful statistical analyses on various historical usage patterns. For example, as we already discussed in section 4.5, the URLs of the Web pages that users are currently viewing will give us information of the current task types. In addition, because Web users have a tendency to choose Websites to visit according to their own previous experiences about the sites, the URLs are also useful for inferring the users' subjective feedback levels on the contents of Web pages if we monitor user activities for a long period. Actually, in the post interviews of the third experiment, the subjects told us that they use different search engines according to their current tasks. For examples, they use Google for careful tasks and Naver – a Korean portal site - for casual tasks. Therefore, we assume that URL information can be used very effectively for the purpose of inferring the user's contexts. The similarity between the contents of current Web pages and contents of previous high-interested Web pages can also be used to infer the interest levels on the current Web pages. Furthermore, the Day Frequency can be used to infer the types of Web pages viewed.

If our system can infer the current contextual factors intelligently, some proactive services can be provided. In figure 10, we present the concept of a data preparation service that we are developing in which unnecessary visit logs and uninterested contents can be filtered out. In addition, if the system can identify a user's current task type correctly, the threshold of the viewing time to find high-interested Web pages can be applied accordingly.

Finally, we should consider individual differences because there may be variances according to user preference, cognitive styles, temperament, and so on.

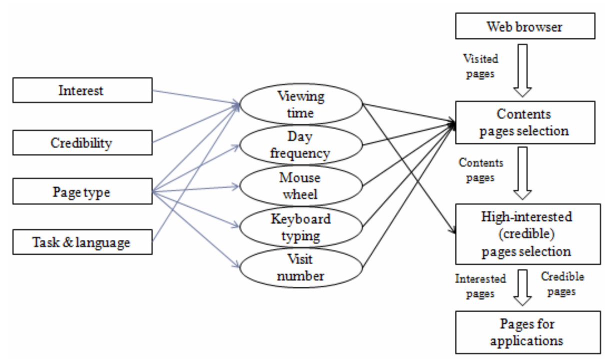


Figure 10. A possible practical solution- the arrows on the left shows their influential relationships and the arrows on the right means that the logs can be used for data preparation tasks

References

Al halabi W. S.; Kubat. M.; and Tapia M. 2007. Time spent on a web page is sufficient to infer a user's interest. In *Proceedings of the IASTED European Conference: internet and multimedia systems and applications*, 41-46, Chamonix, France: ACTA Press.

Badi R.; Bae S.; J. Moore M.; Meintanis K.; Zacchi A; Hsieh H.; Shipman F.; and Marshall C. C. 2006. Recognizing user interest and document value from reading and organizing activities in document triage. In *Proceedings of the 11th international conference on Intelligent user interfaces*, 218-225, Sydney, Australia: ACM.

Borlund, P. and Ingwersen, P. 1997. The Development of a Method for the Evaluation of Interactive Information Retrieval Systems. *Journal of Documentation*, 53(3):225-250.

Byström K. and Järvelin K. 1995. Task complexity affects information seeking and use. *Information Processing and Management*, 31(2):191-213.

Chi E. H.; Pirolli P.; Chen K.; and Pitkow J. 2001. Using information scent to model user information needs and actions and the Web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 490-497, Seattle, Washington, United States: ACM.

Choi J.; Lee G.; and Um Y. 2007. Analysis of Internet Users' Interests Based on Windows GUI Messages. In *Proceedings of the 12th International Conference on Human-Computer Interaction, Lecture Notes in Computer Science*, 4553:881-888.: Springer Berlin / Heidelberg.

Cooley R.; Mobasher B.; and Srivastava J. 1999. Data Preparation for Mining World Wide Web Browsing Patterns. *Knowledge and Information Systems*, 1(1):5-32.

Domenech J. M. and Lorenzo J. 2007. A Tool for Web Usage Mining. In Proceedings *of the 8th International Conference on Intelligent Data Engineering and*

Automated Learning, *Lecture Notes in Computer Science*, 4881:695-704, D.: Springer Berlin / Heidelberg.

Fogg B. J.; Soohoo C.; Da-nielson D. R.; Marable L.; Stanford J.; and Tauber E. R. 2003. How do users evaluate the credibility of Web sites?: a study with over 2,500 participants. In *Proceedings of the 2003 conference on Designing for user experiences*, 1-15, San Francisco, California: ACM.

Fu Y.; Shih M.; Creado M.; and Ju C.. Reorganizing web sites based on user access patterns. 2001. In *Proceedings of the tenth international conference on Information and knowledge management*, 583-585, Atlanta, Georgia, USA,: ACM.

Gauch S.; Speretta M.; Chandramouli A.; and Micarelli A. 2007. User Profiles for Personalized Information Access. The Adaptive Web, *Lecture Notes in Computer Science*, 4321:54-89,: Springer Berlin / Heidelberg.

Hofgesang P. I. 2006. Relevance of Time Spent on Web Pages. In *Proceedings of KDD Workshop on Web Mining and Web Usage Analysis, in conjunction with the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Philadelphia, PA.

Johnson J. D. 2003. On contexts of information seeking. *Information Processing & Management*, 39(5):735-760: Elsevier

Kari J. and Savolainen R. 2007. Relationships between information seeking and context: A qualitative study of Internet searching and the goals of personal development. *Library & Information Science Research*, 29(1):47-69: Elsevier

Kellar M.; Watters C.; Duffy J.; and Shepherd M. 2005. Effect of Task on Time Spent Reading as an Implicit Measure of Interest. In *Proceedings of the American Society for Information Science and Technology*, 41(1):168-175.

Kellar M.; Watters C.; and Shepherd M. 2007. A Field Study Characterizing Web-based Information Seeking Tasks. *Journal of the American Society for Information Science and Technology*, 58(7):999-1018: John Wiley & Sons.

Kelly D. and Belkin N. J. 2004. Display time as implicit feedback: understanding task effects. In *Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, 377-384, Sheffield, United Kingdom: ACM.

Kelly D. and Cool C. 2002. The effects of topic familiarity on information search behavior. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on Digital libraries*, 74-75, Portland, Oregon, USA.

Kelly D. and Teevan J. 2003. Implicit feedback for inferring user preference: a biblio-graphy. *ACM SIGIR Forum* 37(2):18-28. ACM.

Kelly D. 2004. Understanding implicit feedback and document preference: a naturalistic user study. Ph.D. Dissertation, Rutgers University. 2004.

Kim K. and Allen B. 2002. Cognitive and task influences on Web searching behavior. *Journal of the American Society for Information Science and Technology*, 53(2):109-119: John Wiley & Sons.

Liu Z. 2005. Reading behavior in the digital environment. *Journal of Documentation*, 61(6):700-712. Emerald Group Publishing Limited.

Mobasher B.; Cooley R.; and Srivastava J. 2000. Automatic personalization based on Web usage mining. *Communications of the ACM*, 43(8):142-151: ACM.

Nakamichi N.; Shima K.; Sakai M.; and Matsumoto K. 2006. Detecting low usability web pages using quantitative data of users' behavior. In *Proceedings of the 28th international conference on Software engineering*, 569-576, Shanghai, China: ACM.

Rieh S. Y. 2002. Judgement of information quality and cognitive authority in the Web. *Journal of the American Society for Information Science and Technology*, 53(2):145-161: John Wiley & Sons.

Salton G. and McGill M. J. 1986. Introduction to Modern Information Retrieval: McGraw-Hill.

Seo Y. W. and Zhang B. T. 2000. Learning user's preferences by analyzing Web-browsing behaviors. In *Proceedings of the fourth international conference on Autonomous agents*, 381-387, Barcelona, Spain: ACM.

Sonnenwald D. H. 1999. Evolving Perspectives of Human Information Behavior: Contexts, Situations, Social Networks and Information Horizons. Exploring the contexts of information behaviour, 176-190: Taylor Graham Publishing.

Vakkari P. 1999. Task complexity, problem structure and information actions: integrating studies on information seeking and retrieval. *Information processing & management*, 35(6):819-837: Elsevier.

Vakkari P. 2001. A theory of the task-based information retrieval process: a summary and generalisation of a longitudinal study. *Journal of Documentation*, 57(1):44-60: Emerald Group Publishing Limited.

Wang Y. D. and Emurian H. H. 2005. An overview of online trust: Concepts, elements, and implications. *Computers in Human Behavior*, 21(1):105-125: Elsevier.

Wang P.; Hawk W. B.; and Tenopir C. 2000. Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. *Information processing & management*, 36(2):229-251: Elsevier.

Wathen C. N. and Burkell J. 2001. Believe it or not: Factors influencing credibility on the Web. *Journal of the American Society for Information Science and Technology*, 53(2):134-144: John Wiley & Sons.

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*, 21-30, Banff, Alberta, Canada: ACM.