*Journal of Universal Computer Science, vol. 16, no. 16 (2010), 2232-2251*
*submitted: 31/3/10, accepted: 31/7/10, appeared: 28/8/10 © J.UCS*

# Web Context Classification Based on Information Quality Factors

**Jinhyuk Choi**
(Korea Advanced Institute of Science and Technology (KAIST), Korea, Republic
demon@kaist.ac.kr)

**Geehyuk Lee**
(Korea Advanced Institute of Science and Technology (KAIST), Korea, Republic
geehyuk@kaist.ac.kr)

**Junghoon Moon**
(Seoul National University, Korea, Republic
moonj@snu.ac.kr)

**Abstract:** The fact that the World Wide Web is being used for various purposes also implies that users may have various information quality factors to consider according to their current context. In this regards, it is important for Web recommendation services to recognize what quality factors should be considered in current context in order to enhance user satisfaction. We showed that it is necessary to classify Web contexts based on the information quality factors users consider in their minds when they choose websites or Web pages. The results of user interviews showed that there are four quality factors: credibility, recency, popularity, and relevance. From survey data analysis, we recognized that user tasks can be clustered into two groups based on the quality factors that users consider. Finally, the results of log data analysis and performances of our proposed algorithm showed that it is possible to enable Web services to infer the context group. This result implies that context recognition is possible using the limited data that are collected at browser side.

**Keywords:** World Wide Web, Context Classification and Inference, Information Quality Factors, Web Recommendation Services
**Categories:** H.5.0, H.3.5, H.1.2, M.5

## 1 Introduction

As the World Wide Web has become integrated into our routine as a principle form of information media [Rieh, 04], it is typically used not only to search for information but also for purposes such as distraction, entertainment, communication, shopping, learning, and many others. Many portal sites provide information about nearly all aspects of our lives as well as an assortment of issues. News sites let us know of important events in the world in real time. Online shopping sites enable us to buy various products, even without visiting markets. People can encounter various types of leisure content at entertainment sites. Blogs, online community sites and social network services are playing important roles as communication tools.

Information quality consists of various factors - information accuracy, output timeliness, reliability, completeness, relevance, precision, and currency [DeLone, 92].

This means that Web users may have various information quality factors to consider when they choose contents to view. For example, users may consider not only the relevance of the content between their information needs and the content of Web pages they visit, but also the information sources of the content or the author of the content. On the other hand, there may also be a situation in which the time that has been passed since the content was created is considered seriously. Recency may be more important when users choose what content to view at newspaper sites. Users may expect a less serious attitude while they visit different websites for the purpose of distraction or simply to kill time.

However, most context-aware Web services consider a context that focuses only on the role of the Web as an information seeking tool; thus, only the semantic aspects of information quality have been emphasized to deliver relevant contents. For examples, the lists retrieved when using search engines are personalized based on user information profile [Tan, 06]. The most suitable queries are recommended to users [Chirita, 07] to help them seek their target information efficiently. Item-based or content-based recommendation services provide the most relevant content with profiles that are constructed based on users' previous Web visit histories [Gauch, 07]. These services basically attempt to construct user profile and use the profile to infer user information needs. New contents that are semantically similar with the needs are recommended. They have provided good solutions to information overload problems in that they alleviate users' risks to visit a lot of irrelevant Web pages searching for their target information. On the other hand, they tended to recommend similar contents repeatedly so that overspecialization effects have been caused [Sinha, 01].

The primary objective of this study is to classify Web contexts based on the range information quality factors users require when they choose websites or Web pages. User studies are conducted to seek information quality factors and to determine how Web contexts can be categorized according to the required information quality factors. An algorithm that enables Web recommendation services to recognize what information quality factors a user currently require is introduced.

## 2 Problem Definition and the Proposed Approach

The main assumption of this study is that Web users may have various information quality factors (QFs) to consider according to their current purposes in using the Web. Therefore, to enhance user satisfaction, it is necessary for Web recommendation services to infer current user QFs on the fly. The services should select the information to recommend by considering the current QFs. From this perspective, this study is conducted to find answers of several questions as follows:

1. What QFs exist in the actual use of the Web?
2. How Web context can be categorized based on QFs?
3. How context aware recommender system can infer current user QFs and their change trends on the fly?
4. Is it possible for the system to infer current user QFs using limited available Web usage logs at browser side?

| | **User Interview** | **Questionnaire analysis** | **Log data analysis** |
|---|---|---|---|
| **Number of subjects** | 22 | 48 | Lab. Experiment : 20<br>Field study : 12 |
| **Age** | 26 (Average) | | |
| **Gender** | Male: 82% / Female:18% | | |
| **Web usage period** | 10.5 years (Average) | | |
| **Web expertness (self-estimated in 5 point scales)** | 4.09 (Average) | | |

*Table 1: Subjects information*

Many studies have attempted [Johnson, 03; Wang, 00; Kari, 07] to construct Web context model. However, typically only qualitative analyses were conducted by the studies because the purpose was not to develop an intelligent system but to construct a theoretical model. How Web usage patterns reflect the influences of various factors was not studied quantitatively. In this study, quantitative analyses to determine patterns from log data as well as qualitative analyses to identify QFs are conducted because context-aware Web recommendation services should recognize the effects of user QFs on the patterns of log data.

The research method in this study consists of several phases - user interviews, surveys, log data collection and analyses, and context inference algorithm design and evaluation. A qualitative analysis is conducted by way of user interviews. The data from the surveys and log data collection are analyzed quantitatively. The objectives of the first two phases are to seek user QFs during their actual use of the Web and to determine how Web contexts can be categorized according to the QFs. The objectives of last two phases are to determine Web usage patterns that may vary depending on the context and to develop a context inference algorithm using the patterns. The details of each research phase – the number of subjects, question lists for user interviews and surveys, log data collection procedures, and so on - were determined with reference to previous works [Wang, 00; Kelly, 04; Kari, 07; Kellar, 07].

To begin with, user interviews are conducted to collect the practical and typical daily Web usage patterns and to determine the directions of subsequent research based on the collected patterns. In a user interviews, the subjects' verbal expressions concerning the content they seek from the Web are recorded. In addition, the QFs they have in mind when they choose a website to visit and Web pages to view are assessed. In surveys, the scores that the subjects mark regarding how much they consider QFs while they are conducting Web tasks are collected. Analyzing the patterns in the collected scores, Web contexts are classified. In log data collection and analysis, the actual Web usage data collected while the subjects use the Web under the classified contexts are analyzed. The usage data include not only data that are collected while the subjects conduct different Web tasks in the laboratory – in a laboratory experiment – but also the data that are collected in the subjects' own residences for a period of two weeks – a field study. In the context inference algorithm design and evaluation phase, an algorithm is developed to infer the current

Web context intelligently based on the patterns discovered during the collection data. The performance of this algorithm is evaluated using the data collected from the field study. In each phase, undergraduate students and graduate students are recruited as subjects because the study sought to examine experienced Web users who use the Web for various purposes. Table 1 describes the subjects.

## 3 Literature Review

Many studies have investigated the factors considered by Web users when they access Web pages. In one such study [Barry, 98], information seeker relevance criteria were classified into several factors, including a content factor, a personal factor, and a quality factor; and information accuracy, novelty, document reputation, and recency were considered as elements for the factors. In another study [Tombros, 05], text elements, structure elements and quality elements were identified as important document features that users consider while they are conducting practical search tasks. Particularly, scope/depth, authority, recency, and novelty were included as quality elements. Authority was studied as an important factor in a recent study [Rieh, 02]. The elements of credibility factors were found and their influences were measured [Fogg, 03]. In several studies related to online shopping [Li, 04], the responses of other users regarding Web content were emphasized. Perceived playfulness, confirmation to satisfaction, and perceived usefulness were discovered as crucial factors that contribute to users visiting some websites repeatedly [Lin, 05]. From these studies, various factors were found that users have in mind when accessing Web pages.

Despite the fact that there is a variety of factors to be considered by users as they access the contents of Web pages, the research in the field of Web context-aware services has focused mainly on accuracy of recommendations. Web personalization and recommendation services – among the most representative adaptive services in the Web environment – use pre-constructed user profiles to generate recommendations. In these services, user profiles are generally represented in a format of item vectors or what is known as a ‘bag of words’ extracted from the visit histories of the users. As an example, in typical context-aware Web services, users’ Web navigation patterns or query input patterns as they pertain to search engines have been analyzed to determine users’ precise information needs [Chirita, 07; Cooley, 99; Fu, 01; Joachims, 05; Tan, 06]. User interaction patterns with Web pages have been studied to elicit their interests [Badi, 06; Hofgesang, 06; Kellar, 04; Kelly, 04]. User visit histories have been analyzed to build profiles and personalized contents are provided based on profiles [Gauch, 07]. The services choose contents to recommend using various statistical techniques with which semantic distance between profiles and contents can be measured [Brusilovsky, 07]. In fact, because the Web is basically an information seeking tool and because Web users often experience information overload, it is natural to consider that one of the most important contexts that the services have to recognize is the content a user is seeking. However, as users have various factors in mind when assessing Web pages, it is too much of simplification to only consider the semantic aspects of target information to provide context-aware Web recommendation services.

| **Mostly mentioned topics** | **Frequently visited websites** |
|---|---|
| Communication with friends or family members<br>- Daily correspondences<br>- Light conversation<br>- Photo sharing | Cyworld (http://www.cyworld.com/)<br>Private Blogs<br>Me2day (http://www.me2day.net/)<br>Twitter |
| Searching for academic information<br>- Papers, reports, research material<br>- Patents | Google scholar<br>ACM<br>IEEE Explorer<br>The patent office homepage<br>Google patent |
| Pastime, entertainment<br>- TV program<br>- Videos (UCC)<br>- Music<br>- Movies<br>- Sports<br>- Humor<br>- Cartoons | Naver (http://www.naver.com/)<br>Daum (http://www.daum.net/)<br>YouTube |
| News | Naver<br>Daum<br>Newspaper sites |
| Community<br>- BBS in university<br>- Music<br>- Sports<br>- Fashion<br>- Bicycle<br>- Coffee | Naver community<br>Daum community<br>Club websites |
| Programming information<br>- Debugging tips<br>- API documents<br>- Code examples | Google<br>Devpia<br>Source forge<br>Code project |
| E-mail checking | |

*Table 2: Task topics mentioned by subjects in the interview*

There have been several attempts to reflect not only the semantic aspects but also other factors building the formulation of a recommendation list. In one recent study [McNee, 06], the importance of applying many different metrics to build recommendation lists was discussed. In particular, serendipity was considered as one of the most important metrics. In another study [Ziegler, 05], a recommendation system that adopts content diversity in recommendation lists was designed and its usefulness was shown. These studies attempted to reflect variety in their recommendation lists but they did not consider that the variety level can vary dynamically depending on the Web tasks. As found in a study [Choo, 98], there are several behavioural models of information-seeking on the Web. During conditioned viewing, users search for information about selected topics. On the other hand, during undirected viewing, users have no specific information in mind and commonly scan various sources of information broadly. In another study [Kellar, 07], it was suggested that a user's Web tasks consist of such activities as fact finding, information gathering,

| **Tasks** | **Subjects' mentions** | **Related information quality factors** |
|---|---|---|
| Communication with friends or family members | "I search for newly added comments or photos."<br>"I prefer light and funny messages."<br>"I prefer content that shows the writer's character if I don't know it." | Recency<br>Interest<br>Novelty |
| Searching for academic information | "I check the keywords in title or snippets first of all."<br>"I examine conferences or journals by which the paper was published and prefer conferences and journals that have high authority."<br>"I look for the author's current affiliation or position."<br>"I check the articles' publication dates." | Relevance<br>Recency<br>Affiliation<br>Credibility |
| Pastime, entertainment | "I click hyperlinks on which exciting keywords are shown."<br>"I try to find newly published contents."<br>"I choose contents that are related to current big and public issues."<br>"I choose articles that denote a high number of visits or comments."<br>"I am interested in current popular queries that are input by other users." | Interest<br>Recency<br>Popularity |
| News | "I choose mostly some of the headlines that are provided by newspaper sites or portal sites."<br>"I prefer articles that are read by many users."<br>"I look for flash news."<br>"I choose news that is related to current big and public issues." | Recency<br>Popularity |
| Community | "I read new articles or comments."<br>"I choose contents that look interesting."<br>"I choose articles that mark high number of visits or comments." | Recency<br>Interest<br>Popularity |
| Programming information | "I choose articles that include keywords that are related with my information needs on title."<br>"I prefer content that is written by credible authors." | Relevance<br>Credibility |

*Table 3: Information quality factors the subjects considered while conducting the Web tasks*

simple browsing, and making transactions. Among these tasks, fact finding was defined as a task in which users are looking for specific information, and browsing was defined as a task where users are visiting Web pages with no specific goal to see "what's new". This implies that it is necessary for a recommendation service to infer the QFs that a user has in mind – for example, whether the user is searching for relevant information or just scanning new and various contents. It is also necessary to select information to recommend according to the inferred QFs to increase user satisfaction.

## 4 Qualitative Analysis – User Interview

User interviews were conducted face-to-face with 22 subjects. The subjects were asked to describe what tasks they conduct usually, what types of information they seek from the Web in each task, and what QFs they have in mind when they choose websites to visit as much detail as possible. The question list was constructed based on Uden's question list [Uden, 07] to search for QFs based on Activity theory. Each subject came to our laboratory 4 times and was asked to describe different types of task in each time so that the interviews took a total of over 88 hours. The total number of subjects was determined to be over 20 with reference to previous works [Wang, 00; Kari, 07]. Some of the subjects gave answers about their usage patterns based solely on their memories of their online activity, but most gave answers using the Web on computers in our laboratory. The subjects were free to tell us about their Web usage patterns in an open-ended manner; hence, it was possible to obtain actual examples. In this paper, the results of user interview were introduced briefly due to constraints of space.

They described various types of Web tasks as shown in Table 2. The task topics can be categorized into intimate communications, research material searches, consumption of entertainment information, visits to news sites, visits to community sites, searches for programming-related information, e-mail checking, and others. Other than the topics that are listed in Table 2, a small number of subjects told us that they look up new words in a dictionary, look for restaurants, use online banking sites, and complete other tasks.

Their preferred websites differ according to the objectives of their task and/or topic. In other words, they visit websites that are specific to each task. For example, in the case of portal sites, they do not use several sites but visit a preferred site repeatedly. They told us that it is not necessary to visit several portal sites to obtain new content because the sites provide nearly the same content equally. In the case of newspaper sites, they choose a site based on their political preference. They tended to use Google to seek international information that requires credibility, but use Korean

| | **Task topics** |
|---|---|
| Task1 | Entertainment –related (sports, movie star, music, humor, etc.) |
| Task2 | News (topics of the day – politics, economics, gossips, etc.) |
| Task3 | Livelihood or shopping related (health, personal knowledge, new product, etc.) |
| Task4 | Research-related (paper, reports, homework, etc.) |
| Task5 | Programming-related (sample code, API information, etc.) |
| Task6 | Social network-related (intimate communications, photo and video sharing, etc.) |

*Table 4: The topics of representative tasks*

portal sites – for examples, Naver (http://www.naver.com/) and Daum (http://www.daum.net/) - to obtain informal or unofficial information.

Table 3 shows examples of the comments that were collected within the interviews. From these comments were extracted the QFs considered by users while they conducted various Web tasks. There were four main QFs to consider: credibility, recency, popularity, and relevance. Among the four factors, credibility is a feeling about how believable the content of a website or a Web page is. Credibility is assessed mainly based on the sources of the content, the reputation and the authority of the website or the author, and similar factors. Recency is mainly assessed using the day and time when the content was written, and popularity is mainly assessed based on the number of visits by other users and the number of replies or responses regarding the content. Relevance is assessed using the keywords in the titles of the content. Some of the QFs – affiliation and interest in Table 3 – were not included in the main QFs because they were regarded as secondary attributes by which credibility and relevance can be assessed. The main QFs that we found is a subset of QFs that can be found in general information seeking behaviours [DeLone, 92]. Some factors that were included in the QFs in general information seeking behaviours were not found in our interviews. To the best of our beliefs, distinctive characteristics of the Web environments made the differences.

# 5 Quantitative Analysis

## 5.1 Survey Data Analysis

Among the Web tasks, six tasks were selected as representative tasks. The six tasks were selected because the tasks were mentioned by nearly all of the subjects in the user interview. The subjects rated how much the four QFs they consider while they conduct the selected Web tasks. The topics of the representative tasks are listed in Table 4. For credibility, recency, and popularity, the subjects expressed their rates on

| **Topic examples** | **Low contents diversity (high-relevance)** | **High contents diversity (low-relevance)** |
|---|---|---|
| Media-law related news article | Other media-law related news articles | Other news articles related to politics- |
| Papers that were published by Dr. Kim, who studies Computer science | Other papers that have similar topics | Other papers in Computer science |
| Portal site content about Wonder Girls (Korean idol stars) | Other content related to Wonder Girls | Other content in the entertainment world |
| Portal site content about Chan Ho Park | Other content related to Chan Ho Park | Other content in sports |
| MySQL programming-related content | Other MySQL programming-related content | Other content in the field of database |

*Table 5: Examples in questionnaire sheet to help the subjects understand the concepts of contents diversity*

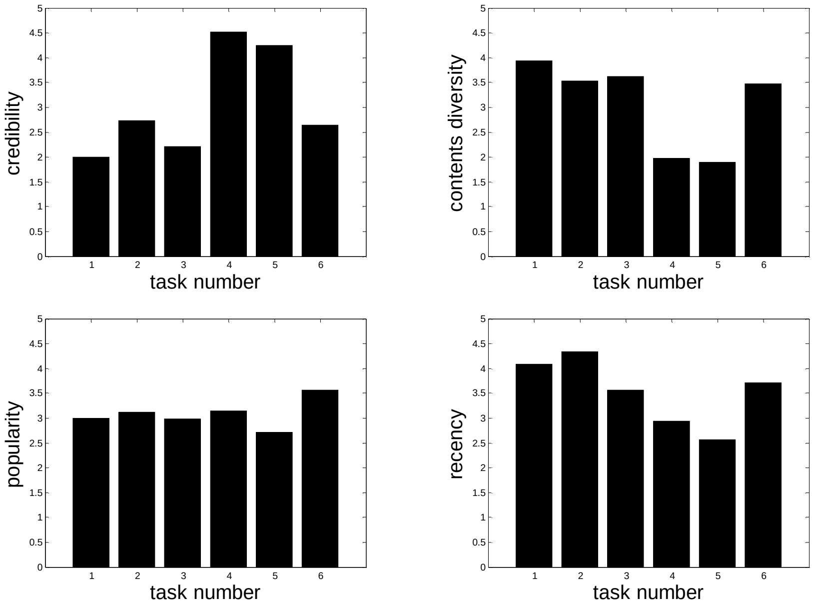
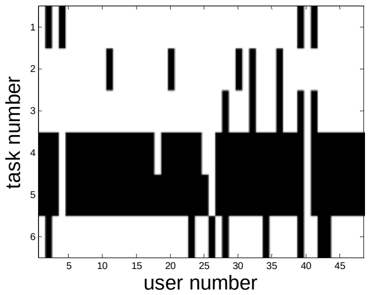


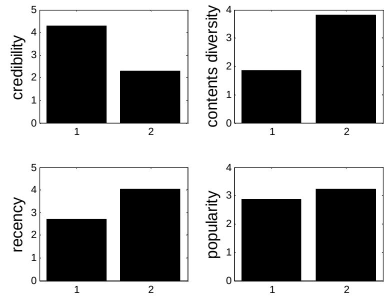

*Figure 1: the mean scores of expressed rates for QFs*

five-point scales. For relevance, the subjects expressed their rates considering the level of content diversity they prefer when choosing new Web pages as they conduct the tasks. The subjects were allowed to see some examples (Table 5) before selecting answers to the relevance questions to help them understand the concept of content diversity. The subjects rated the level of acceptable content diversity on a scale from 1 – a low level of acceptable diversity (only high relevance acceptable) to 5, a high level of acceptable diversity (low relevance also acceptable).

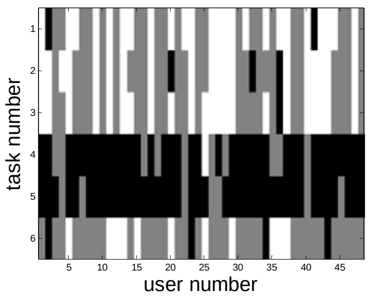
Figure 1 shows how much the QFs were considered according to the task. A one-way ANOVA test was conducted to explore the differences of scores. Statistically significant differences among the tasks ($p < 0.01$) were found in the mean scores of credibility, recency, and relevance, but not in popularity. A multiple comparison test was conducted to determine which pairs of means were significantly different. For credibility, Task 4 and Task 5 showed higher scores than the other tasks, and the difference between high scoring task groups (Task 4 and Task 5) and low scoring task groups (Task 1, Task 2, Task 3, and Task 6) was significant. For recency, Task 2 showed the highest score. On the other hand, Task 5 showed the lowest score for recency. The recency scores also showed significant differences between highly scored task groups and low scoring task groups. In the case of content diversity, Task 4 and Task 5 showed lower scores than the other tasks, and the difference between the low scoring task groups (Task 4 and Task 5) and highly scored task groups (Task 1, Task 2, Task 3, and Task 6) was significant. Essentially, the content diversity scores showed patterns that were opposite to those of credibility. In contrast to the other factors, the popularity scores among the tasks showed no significant differences.



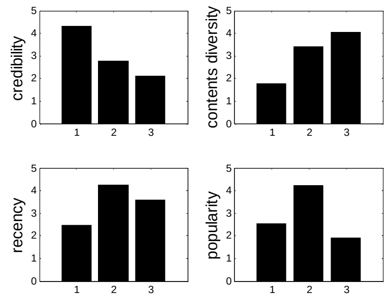

(a) K =2 (black – cluster 1 / while – cluster 2)




(b) K =2 (x- axis: cluster number / y-axis: mean of rates)




(c) K =3 (black – cluster 1 / gray – cluster 2 / white – cluster 3)




(d) K = 3 (x- axis: cluster number / y-axis: mean of rates)

*Figure 2: Cluster analysis results*

A cluster analysis was conducted to determine whether there were noticeable clustering patterns in the distribution of users' rates for the QFs. The tasks were expressed as four-dimensional vectors in which each dimension represented a QF that a user rated. 48 subjects rated 6 tasks for a total of 288 vectors. A k-means clustering algorithm was applied and vector similarities were measured based on the Euclidian distance. Figure 2 and Table 6 show the results of the clustering analysis.

| | **Two-cluster** | | **Three-cluster** | | |
|---|---|---|---|---|---|
| | **Cluster #1** | **Cluster #2** | **Cluster #1** | **Cluster #2** | **Cluster #3** |
| Task 1 | 4 | 44 | 3 | 23 | 22 |
| Task 2 | 5 | 43 | 2 | 26 | 20 |
| Task 3 | 5 | 43 | 2 | 23 | 23 |
| Task 4 | 43 | 5 | 43 | 4 | 1 |
| Task 5 | 45 | 3 | 44 | 3 | 1 |
| Task 6 | 8 | 40 | 8 | 29 | 11 |

*Table 6: Number of task vectors that were assigned to clusters*

K=2

In the results of the two-cluster clustering, the subjects' rates for the QFs clearly differed among the tasks. As shown in Figure 2 - (a) and Table 6, the vectors of task 4 and task 5 were mostly assigned to the same cluster (cluster 1) while the other vectors were assigned the other cluster (cluster 2). The credibility rates of the tasks in cluster 1 were significantly higher than the tasks in cluster 2. On the other hand, the diversity rate, the recency rate, and the popularity rate of tasks in cluster 1 were significantly lower than the tasks in cluster 2 (Figure 2-(b)). This indicates that the tasks can be divided into two groups. The first group contains the tasks that require relatively high credibility and relevancy. The second group contains the tasks that require relatively high recency and popularity.

K=3

For the three-cluster clustering, as shown in Figure 2 - (c) and Table 6, the vectors of task 4 and task 5 continued to be assigned mostly to the same cluster (cluster 1). As with the two-cluster clustering results, the credibility rates of the tasks in cluster 1 were significantly higher than the tasks in other clusters while the diversity rate of tasks in cluster 1 were significantly lower than the tasks in other clusters. However, the vectors of task 1, task 2, task 3, and task 6 were assigned evenly in both cluster 2 and cluster 3. These tasks could not be clearly clustered by the rates of QFs.

The results show that the tasks can be clustered into two groups more clearly than three groups based on the QFs and that the subjects have different QFs according to the groups. When they are viewing research-related content or looking for programming-related information, task 4 and task 5, they think much of credibility and relevance. In contrast, when they are viewing entertainment content or a newspaper, they take recency and popularity more seriously than credibility or relevance. The two groups were termed the Careful Web context group and the Causal Web context group, respectively. The characteristics of the groups are presented in Table 7. It is important for Web service developers to classify the tasks into groups according to the rate of each QF. For example, Web pages that contain credible content and relevant content considering a user's current target information should be included mainly in the recommendation list when users are in the Careful

| | **Careful Web context group** | **Casual Web context group** |
|---|---|---|
| Representative topics | Research-related<br>Programming-related | Entertainment -related<br>News<br>Livelihood or product-related<br>Social network-related |
| Web usage patterns | Clear target information<br>Active search mostly using search engines or search interfaces on websites | Non-clear target information<br>Passive selection among the content provided by websites |
| QFs | Credibility and relevance | Recency and popularity |

*Table 7: The characteristics of the context groups*

Web context. In contrast, it is better to consider the recency and popularity rather than credibility or relevance when users are in the Casual Web context.

### 5.2 Log Data Collection and Analysis

The first item a Web service should concern itself with is recognizing the current context intelligently to provide a context-aware Web service. One important factor to consider is that the context recognition process should be performed at browser side because various contextual situations can only be observed when users are visiting various websites to conduct different types of tasks. This implies that only limited log data that can be collected at browser side can be used to recognize the context. Thus, a browser-monitoring module (BMM) that runs behind Internet Explorer without any modification to the browser was developed. A BMM is a type of monitoring software developed to collect the visited URLs, the visit time and the dwell time on each Web page while users used the Web. After Web searching, using a feedback window, users can review the visited Web pages and choose a radio button that represents the context group under which they visited the page. If the user does not want to answer questions regarding a Web page, they can easily remove the record.

The log data that was collected at browser side were analyzed to discover noticeable patterns that could be used to identify the context group. A laboratory experiment was conducted in which 20 subjects participated, and a field study was done in which 12 subjects volunteered. In the laboratory experiment, the subjects came to our laboratory and completed two tasks with their own topics – one task under the Careful Web context and the other task under the Casual Web context – using desktop computers onto which the BMM was installed. The subjects were allowed to conduct the tasks as they usually would for two hours. To obtain appropriate data, the subjects were not told that some activities would be measured while they read the Web pages. As soon as the subjects finished their tasks, they reviewed each Web page that they visited and recorded the context groups under

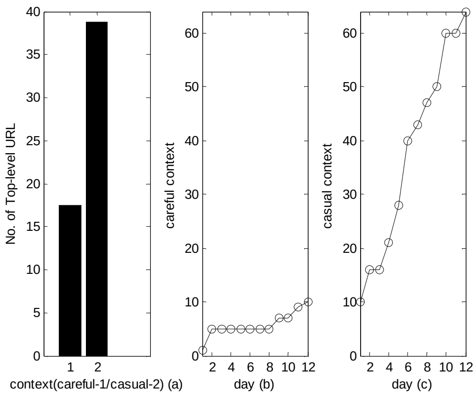



*Figure 3: The average number of URLs in each task (left), the increasing rate of the average number of URLs in the careful context (middle) and in the casual context (right)*

| user No. | Context 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 8: The proportion of context separable URLs*

which they visited the page. In the field study, the subjects installed the BMM on their computer and conducted Web tasks in their own residence for two weeks. The subjects reviewed visited Web pages at least once a day and labelled the context group under which each Web page was visited. The details of this research phase were determined with reference to related previous works [Kelly, 04; Kellar, 07]. During the field study, the subjects visited over 30,000 Web pages. The Web visit patterns that were observed according to the context group are summarized below.

(1) Over 90% of the visited top-level URLs were separable into the context groups. In other words, 90% of visited top-level URLs belonged to a specific context group (Table 8).

(2) The number of top-level URLs visited by users in the Casual Web context group was much higher than that in the Careful Web context group (Figure 3-(a)).

(3) The number of visited top-level URLs during Casual Web context group tasks increased much more compared to Careful Web context group tasks (Figure 3-(b), (c)).

(4) The mean duration of careful tasks was 1.37 minutes (std = 3.07 min.), whereas the mean duration of casual tasks was 3.66 minutes (std = 5.57 min.) when assigning two successive visits in the same session if the difference in the visit time between the visits is not greater than 5 minutes.

Among these observed patterns, the first and second patterns were observed not only in the results of the laboratory experiment but also in the field study. The third and forth patterns were observed in the results of the field study that was conducted over a long period. The first pattern shows that users have their own URL lists that are specific to their current context because they may use the Web based on their previous individual experiences on the Web. The second and third patterns indicate that the subjects showed the patterns of a navigator under the careful context – they were more likely to revisit domains - but showed the patterns of an explorer under the casual context – they tended to branch frequently and visit many new domains [White, 07].

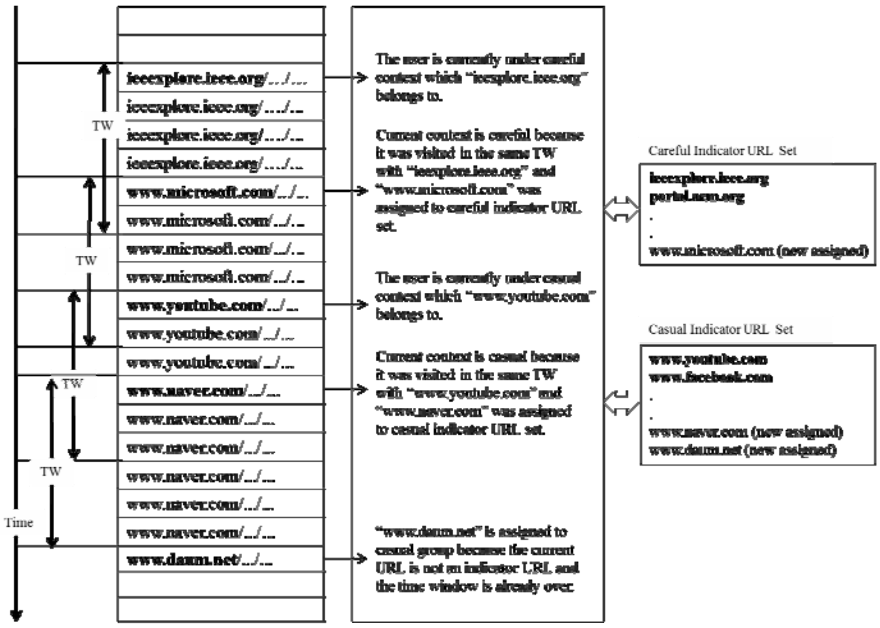



*Figure 4: Concept of context classification algorithm*

## 6 Context Inference Algorithm Design and Evaluation

Based on the patterns discovered thus far, some points were gathered to consider for the development of an intelligent context recognition algorithm.

(1) We may recognize user's current context easily by checking the top-level URLs.
(2) As time passes, newly visited top-level URLs should be added continuously. Therefore, the context group to which a new top-level URL belongs should be identified as soon as possible.

A simple algorithm was developed that uses some of frequently visited URLs as indicators for context groups. It was assumed that the context groups to which the indicator URLs belong to are known to the algorithm in advance. It was also assumed that the unknown URLs that are visited shortly after visiting an indicator URL could be classified into the same groups as the indicator URL. To do this, it was necessary to set a time window (TW) size to decide whether two URLs – indicator URL and the unknown URL - are in the same context group. New unknown URLs were assigned to the context group of the indicator URL that was visited most recently in the same TW. The concepts of the algorithm are presented in Figure 4 and Figure 5. First, when a user visits a Web page, the algorithm checks whether or not the URL is in the indicator URL set. If the URL is one of the indicator URLs, the algorithm assumes that the user is currently in the context group to which the indicator URL belongs. If

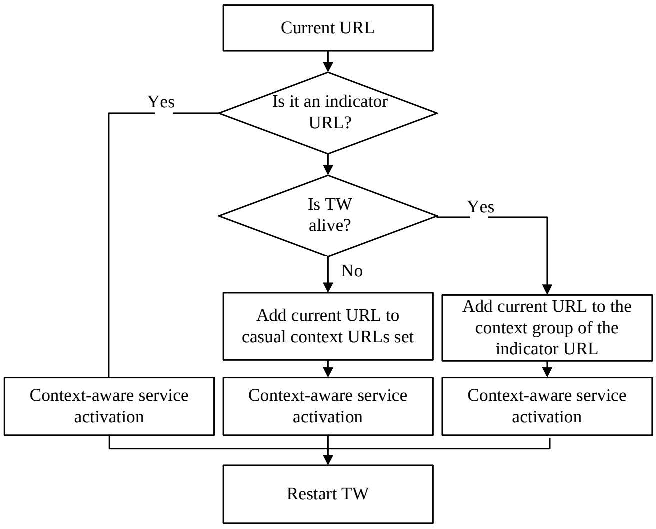



*Figure 5: Context classification algorithm*

the current URL is not an indicator URL, the algorithm checks whether or not the TW is over. If the TW is not over, the current URL is assigned to the context group of the indicator URL that is visited in the same TW. Figure 4 also shows a case in which the current URL is not an indicator URL and where the time window is already over. In this case, the group is set to casual because the total number of visited URLs in the casual group is higher than the total number of URLs in the careful group. A newly assigned URL is also used as an indicator.

Here is an expected situation. A user visits IEEE Explore to search for some materials to help with his homework. As IEEE Explore is an indicator of careful context, the context-aware Web service recommends Web pages that contain content related to what the user is currently seeking. After a few minutes, when the user visits the Microsoft homepage, which is not an indicator URL, to find information about his research topics, the algorithm recognize that the URL should be classified into the careful context group because the homepage is visited before the TW is over. After a while, the user visits YouTube, which was previously classified into the casual group, to watch a video of his friend. The current context is then changed to casual. After he is finished watching the video, the user visits the Naver – one of the most popular portal sites in Korea - website. The algorithm recognizes that the TW is not yet over. Therefore, Naver is classified into the casual group following the context group to

| **User number** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Best TW (min.)** | 1 | 2 | 3 | 1 | 11 | 1 | 3 | 8 | 1 | 2 | 4 | 2 |
| **Best accuracy (%)** | 98 | 80 | 84 | 86 | 78 | 92 | 87 | 93 | 82 | 80 | 94 | 90 |
| **Accuracy (%) (TW=1 min.)** | 98 | 80 | 83 | 86 | 77 | 92 | 86 | 91 | 82 | 80 | 94 | 88 |

*Table 9: Context classification accuracy*

which YouTube belongs. Under this context, the service may recommend Web pages that contain recently published content without much consideration of its credibility or relevance.

The algorithm was evaluated using log data that were collected in the field study. The indicator URLs were collected based on the subjects' feedbacks – from the most frequently visited URLs – and thus the indicator URL lists were different for different subjects. The algorithm used individual indicator URL lists to recognize the current context. Although the recognition performance will improve as the algorithm uses a greater number of indicator URLs, the number of indicator URLs in the Careful Web context group was limited to 5 and the number of indicator URLs in Casual Web context group was limited to 10. The number of indicator URLs was set differently in consideration of the second previously observed pattern in section 5.2: "The number of URLs that users visited in the Casual Web context group was much higher than that in the Careful Web context group." Experiments were conducted to determine the best TW value, as the performance of the algorithm will vary according to the TW value. The results obtained when the TW was set to a fixed value were also analyzed because it was not considered to be easy to determine the best TW values in real time in practical situations.

Table 9 shows the best TW values for each user and algorithm performance when the TW was set to the best value. The performance when the TW was set to a fixed value (1 min - shorter than the mean duration of the tasks in the careful context group) is also presented in Table 9. As the data show, the overall performance was beyond our expectations in consideration of the simplicity of the algorithm. There was no large performance decline, even when the TW was set to a fixed value for convenience.

These results show that it is possible to enable Web services to infer the context group under which a user is visiting new websites or new Web pages if a few context-indicator URLs are known in advance. This result also implies that context-aware Web services that can provide intelligent services to Web users can be developed because context recognition is possible even when using the limited data that can be collected at browser side. The proposed algorithm appears similar with attempts made

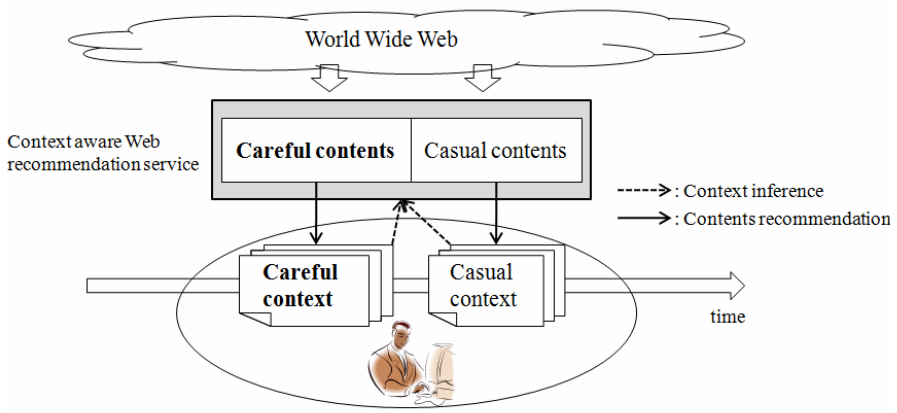



*Figure 6: Concept of context aware Web recommendation services*

in an earlier study [Kellar, 06] using a machine-learning technique. However, the present algorithm is easier and more straightforward to apply in that it simply uses the visit history and involves neither building a statistical model that requires numerous features in its training data nor considerable computational resources.

## 7 Discussion and Conclusion

Context-aware Web recommendation services should recognize a user's current context as early as possible because the user QFs vary according to the context. The services will be able to apply an adequate recommendation policy based on the recognized context.

In this study, two Web context groups - the Careful Web context and the Casual Web context - were found based on the results of user interviews and survey data analyses. In the Careful Web context, users have clear target information to seek and want to find relevant and credible content. In the Casual Web context, users want to view new and popular content even without clear target information. An algorithm was developed to recognize the current context using the log data that was collected at browser side. The algorithm was designed based on the fact that the current context could be recognized without difficulties because visited URLs were separable according to the context group.

The solutions for information overload problems may cause overspecialization problem in that they tend to recommend similar contents repeatedly. On the other hand, the solutions for overspecialization problems may also cause information overload problems because they tend to adopt content diversity in recommendation lists. It means that it is not easy to solve both an information overload problem and an overspecialization problem simultaneously with one single solution. To alleviate user's efforts to locate relevant content, various semantic based recommendation services have been introduced. Speculating that users want to find relevant content mostly while they seek their target information in the careful context, such services are suited to the careful context. On the other hand, several methods to diversify a recommendation list have been discussed to provide users chances to view various new contents. The results of the present study suggest that the methods are applicable, especially for the casual context. Hence, the proposed context recognition processes can address both the information overload problem and the overspecialization problem simultaneously in that it enables recommendation services to apply different recommendation policies according to recognized contexts.

Figure 6 shows the concepts of Web recommendation services that can be developed based on the findings of this study. Context-aware Web recommendation services can apply different criteria for choosing contents to recommend. When a user is in the careful context, the service should recognize the user's information need quickly or consider their profile to find relevant content. In addition, the service should evaluate the credibility of the content using various criteria [Fogg, 03] prior to providing the content to the user. When a user is in the casual context, the service should find how recently the content was published and how many users have viewed the content while also equally inferring what types of content the user has preferred thus far. This implies that recommending recent and popular content is preferred to recommending relevant content.

Although the systemic procedures were painstakingly tasked in this research, it has several limitations. The findings should be verified further based on additional usage data collected from users with various levels of Web experience, as this study recruited only experienced Web users as subjects. Additional example tasks should be collected and studied, as it is also possible that additional context groups can be classified. In this study, several examples were proffered to allow the subject to understand the concept of content diversity, as some of the subjects reported that they could not grasp the exact meaning of this concept. The examples were required but they also may have brought forth a limited understanding in some users, as there is no clear agreement about a method with which content diversity can be measured equally when using various topics.

Further research should be performed to overcome these limitations. Work continues on the development of an advanced algorithm with which context can be recognized even without indicator URLs. To ensure the usefulness of this algorithm, it is necessary to analyze user responses when context-aware Web services that are designed based on the present algorithm are provided. Moreover, according to the tendencies of mobile Web services, studies of the mobile Web context, which may be more complicated and uncertain, are essential. The authors are planning to apply research procedures similar to those used in this study for the classification of different mobile contexts.

**Acknowledgements**

The authors would like to express appreciation to Youmin Kim and Sangyong Gil who helped the authors as interviewers. The authors would also like to express appreciation to graduate and undergraduate students who participated in our research phases as subjects.

## References

[Badi, 06] Badi, R., Bae, S., Moore, J. M., Meintanis, K., Zacchi, A., Hsieh, H.: Recognizing user interest and document value from reading and organizing activities in document triage, Proceedings of the 11th international conference on Intelligent user interfaces, 218-225, Sydney, Australia, 2006

[Barry, 98] Barry, C. L.: Document representations and clues to document relevance. Journal of the American Society for Information Science and Technology, 49(14), 1293-1303, 1998

[Brusilovsky, 07] Brusilovsky, P., Kobsa, A., Nejdl, W. (Eds.).: The Adaptive Web, Methods and Strategies of Web Personalization (Vol. 4321): Springer. 2007

[Chirita, 07] Chirita, P.-A., Firan, C. S., Nejdl, W.: Personalized query expansion for the web, Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, 7-14, Amsterdam, The Netherlands, 2007

[Choo, 98] Choo C. W., Detlor, B., TurnbullD.: A Behavioral Model of Information Seeking on the Web - Preliminary Results of a Study of How Managers and IT Specialists Use the Web, 1998 ASIS Annual Meeting Contributed Paper

[Cooley, 99] Cooley, R., Mobasher, B., Srivastava, J.: Data Preparation for Mining World Wide Web Browsing Patterns. Knowledge and Information Systems, 1(1), 5-32, 1999

[DeLone, 92] DeLone, W. H., McLean, E. R.: Information Systems Success: The Quest for the Dependent Variable. Information Systems Research, Vol. 3 Issue 1, March 1992

[Fogg, 03] Fogg, B. J., Soohoo, C., Danielson, D. R., Marable, L., Stanford, J., Tauber, E. R.: How do users evaluate the credibility of Web sites?: a study with over 2,500 participants, Proceedings of the 2003 conference on Designing for user experiences, 1-15, San Francisco, California, 2003

[Fu, 01] Fu, Y., Creado, M., & Ju, C.: Reorganizing web sites based on user access patterns, Proceedings of the tenth international conference on Information and knowledge management (pp. 583-585). Atlanta, Georgia, USA, 2001

[Gauch, 07] Gauch S.; Speretta M.; Chandramouli A.; and Micarelli A.: User Profiles for Personalized Information Access. The Adaptive Web, Lecture Notes in Computer Science, 4321:54-89,: Springer Berlin / Heidelberg. 2007

[Hofgesang, 06] Hofgesang, P. I.: Workshop on Web Mining and Web Usage Analysis. The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006), 2006

[Joachims, 05] Joachims, T., Granka, L., Pan, B., Hembrooke, H., Gay, G.: Accurately interpreting clickthrough data as implicit feedback, Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval, 154-161, Salvador, Brazil, 2005

[Johnson, 03] Johnson, J. D.: On contexts of information seeking. Information Processing and Management. 39, 5 (Sep. 2003), 735-760, 2003

[Kari, 07] Kari J., Savolainen R.: 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, 2007

[Kellar, 04] Kellar, M., Watters, C., Duffy, J., Shepherd, M.: Effect of task on time spent reading as an implicit measure of interest. Proceedings of the American Society for Information Science and Technology, 41(1), 168-175, 2004

[Kellar, 06] Kellar, M., Watters, C.: Using web browser interactions to predict task, Proceedings of the 15th international conference on World Wide Web, 843-844, Edinburgh, Scotland, 2006

[Kellar, 07] Kellar, M., Watters, C., Shepherd, M.: A field study characterizing Web-based information-seeking tasks. Journal of the American Society for Information Science and Technology, 58(7), 999-1018. 2007

[Kelly, 04] Kelly, D., Belkin, N. J.: Display time as implicit feedback: understanding task effects, Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval, 377-384, Sheffield, United Kingdom, 2004

[Li, 04] Li, X., Ling, L.: PeerTrust: supporting reputation-based trust for peer-to-peer electronic communities, IEEE Transactions on Knowledge and Data Engineering, 16(7), 843-857, 2004

[Lin, 05] Lin, C. S., Wu, S., Tsai, R. J.: Integrating perceived playfulness into expectation-confirmation model for web portal context. Information Management, 42(5), 683-693, 2005

[McNee, 06] McNee, S. M., Riedl, J., Konstan, J. A.: Being accurate is not enough: how accuracy metrics have hurt recommender systems, CHI '06 extended abstracts on Human factors in computing systems, 1097-1101, Montreal, Quibec, Canada, 2006

[Rieh, 02] Rieh, S. Y.: Judgement of information quality and cognitive authority in the Web. Journal of the American Society for Information Science and Technology, 53(2), 145-161, 2002

[Rieh, 04] Rieh, S. Y.: On the web at home: information seeking and web searching in the home environment. Journal of the American Society for Information Science and Technology, 55(8), 743-753, 2004

[Sinha, 01] Sinha, R., Swearingen, K.: Comparing recommendations made by online systems and friends. Paper presented at the In Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries. 2001

[Tan, 06] Tan, B., Shen, X., & Zhai, C.: Mining long-term search history to improve search accuracy, Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining, 718-723, Philadelphia, PA, USA, 2006

[Tombros, 05] Tombros, A., Ruthven, I., Jose, J. M.: How users assess web pages for information seeking. Journal of the American Society for Information Science and Technology, 56(4), 327-344, 2005

[Uden, 07] Uden, L.: Activity theory for designing mobile learning, International Journal of Mobile Learning and Organisation, 1(1), 81-102, 2007

[Wang, 00] Wang P., Hawk W. B., Tenopir C.: Users' interaction with World Wide Web resources: an exploratory study using a holistic approach. Information processing & management, 36(2):229-251: Elsevier. 2000

[White, 07] White, R. W., Drucker, S. M.: Investigating behavioral variability in web search, Proceedings of the 16th international conference on World Wide Web, 21-30, Banff, Alberta, Canada, 2007

[Ziegler, 05] Ziegler, C.-N., McNee, S. M., Konstan, J. A., Lausen, G.: Improving recommendation lists through topic diversification, Proceedings of the 14th international conference on World Wide Web, 22-32, Chiba, Japan, 2005