

Query Cluster: A Method for Web Search Behavior

Jinyoung Kim
 LG Electronics
 Seoul, Korea
 e-mail: amjinyoung.kim@lge.com

Moonsung Kim
 Graduate School of Convergence
 Science and Technology
 Seoul National University
 Suwon, Korea
 e-mail: amenra@snu.ac.kr

Joongseek Lee
 Graduate School of Convergence
 Science and Technology
 Seoul National University
 Suwon, Korea
 e-mail: joonlee8@snu.ac.kr

Abstract— There have been intensive research on user web search behavior since the late 1990s. Previous researchers collected data from search engines and analyzed explicit data (queries) to understand the characteristics of the user’s search process, while other researchers analyzed the data of recruited subjects under experimental settings to understand the behavioral patterns in web usage. Although these researches provided an understanding of what users are searching for and how they are searching, both approaches did not provide rich user contexts that capture the reason why users are motivated to search, how long users’ tasks (session) last, and other factors affecting user’s search behavior. In this paper, we propose ‘Clustered Query’ as the unit of analysis in web search behavior studies. We found that users make their own Clustered-queries that yield better overview on their web search pattern, yet detailed individual web traces intact. The methodology consists of three phases and *Log Catcher*, *Query Cluster*, *Monitoring tool*, and *Retrospective Interview* technique are used in each phase. At the end of this paper, we also illustrate the process of the pilot and main study where the methodology is modified and validated.

Keywords—Web search behavior; Methodology; User Intent; User Context

I. INTRODUCTION

As stated in the 2005 Pew Internet Report, ‘Web has become the new normal’ in the way of modern life [1]. Web and information retrieval has become the dominant issue in the field of information studies ever since. With the development of web and mobile, there also have been changes of human information behavior. The strategic redesigning of web search services such as Google [2] and NAVER [3] has brought major changes from the way we recognize the needs of information to the way we engage information seeking behavior. Search assistance features such as ‘real-time issues’ and ‘related keywords recommendation’ have opened up new ways of searching by generating user’s needs or by providing shortcuts to reach the information a user wants. The Web serves users’ daily information behavior, and the mobile platform is accelerating this phenomenon. Users no longer seek for information just for their jobs, tasks, or expertise but also for everyday curiosity and fun. Even more, they do not need to seek for information as the information comes to the users.

However, previous researchers have focused on the framework that illustrates user’s information behavior and the task-related information needs and process. Although these studies have contributed to the information behavior studies, several constraints are also perceived such as a lack of empirical studies supporting the framework and a lack of user data in natural settings. The studies mostly relied on the qualitative research methodology such as in-depth-interviews to acquire user data. Other researches concentrated on the analysis of users’ daily web usage and search patterns using quantitative data collected through search engine logs or customized tools. These researches are restricted to understand user’s context as they collected and analyzed a ‘series of queries’ that random users typed.

In this paper, we describe a methodology to capture the user’s usual web search behavior and the context of the web search behavior. The methodology allows researchers to collect data from users’ web activity logs in natural settings and accumulations of context to affect the web user’s search behavior. The rest of the paper is organized as follows: In Section 2, we reviewed the previous studies of web search log. The method, Query Cluster, and the refinement of method are introduced in Section 3 and Section 4, respectively. We concluded Section 5 with the discussions and future steps of the study.

II. RELATED WORKS

Web log data allowed researchers to track back user’s information behavior rather than to assume with user’s recollection or diary data. According to Jansen and Spink [4], web-searching studies can be categorized into three methodologies: (1) transaction-log analysis, (2) experimental setting analysis, and (3) issues related to web searching. In this paper, we focus studies on the transaction-log analysis and on the experimental setting analysis.

Transaction-log analysis web-searching studies are one of the major streams that analyze data acquired from search engines to understand the characteristics of web searching behavior. These researches are meaningful as most web users gather to search the engine/portal looking for new information. Researchers extracted the characteristics of web searching by investigating the frequency of query occurrence, the average length of query, the typical query session, or the relevance among queries to improve current web search engine [4][5].

One approach of transaction-log studies is to investigate tactics or strategies in user's web searching. Silverstein et al. and Jansen examined query characteristics and correlation of query logs from the Alta-Vista search transaction-logs [5][6]. Different approaches have been applied to cluster and classify search queries. Ross and Wolfram analyzed the co-occurrence of query terms among the Excite search engine queries [7]. They presented a hierarchical cluster analysis comprising the topics. Shi and Yang also developed a method to identify related queries by extracting and segmenting query sessions and mining association rules from a Taiwanese search engine [8]. While these studies mostly focused on extracting topics of query terms, Rose and Levinson were concerned with understanding the users' intrinsic goals of user searches [9]. They characterized the user search goals – Navigational, Informational, and Resource – that are derived from Broder's 'Taxonomy of Web Search', and manually classified the searching queries of the three goals [10].

Transaction-log studies have strength as they deal with a large number of data of random users, and less likely to be affected by trends [5]. However, it is hard to observe the behavioral pattern of a user and to understand the user context with anonymously collected data. Researchers have to rely on the log data that shows when users search and what they search for, and cannot report in a user-centered manner because of the lack of contextual information [11]. It also has limitation that the analysis may reflect the characteristics of the search engine.

Experimental setting studies, on the other hand, are to analyze data acquired from the customized tools installed on the participant's computer or using the web browser. Participants are recruited for the experiment and their web search pattern is analyzed while the transaction-log studies mainly focus on the analysis of obtained data. These researches cover topics from the characteristics of interaction during information seeking to the context of information seeking.

Choo et al. observed the web seeking behavior of 34 knowledge workers to find out their information needs and information seeking preferences [12]. They extracted the significant episodes during web usage through in-depth-interviews. A customized tool, WebTracker, collected participants' web log data of URL calls/requests, browser menu selections (i.e., reload, back, and forward) and the collected data was used as the background information for the interview. They identified 61 significant episodes of information seeking and categorized them into 4 complementary modes of information seeking.

Sellen et al. studied how and why knowledge workers use the web with a methodology that combined diaries and interviews. They interviewed 24 workers about their web search history with web history references written on the worker's personal computer. Participants were asked to tell a story of their searching activities and to rate their web activities with respect to the success/failure, significance of

the activity, and time spent on the activity [13].

Kelly proposed a method for collecting the user data about information seeking contexts and behaviors in natural environments [14]. Seven PhD students used laptops equipped with a client-side logger. The students reported 5 variables – endurance, frequency, stage, persistence, and familiarity- related to the tasks and topics of their web seeking behavior, and usefulness ratings and confidence of the document. More details were obtained through the exit interview at the end of the research.

Kellar et al. also examined how users interact with their web browsers during information-seeking tasks [15]. 21 students installed a custom-built web browser that collected visited websites and browser menu logs. Students reported their own browsing histories in task types -fact finding, information gathering, just browsing, transactions, and others- and task descriptions through electronic diaries or real-time reports. Experimental setting analysis usually uses a combined methodology to obtain qualitative data and quantitative log. A small group of participants is recruited for the research and the customized tool collects user's web log and interview follows. Although the experimental setting analysis provides qualitative data of user context, it still has restrictions of small data sets. Also participants sometimes forgot about the past research behavior as interviews or clustering assignment are delayed [14] and researchers missed the details of user context as they focused on the browser controlling behavior [15].

The purpose of this study is to develop a method to overcome limitations of prior studies. We used 'Clustered Query' as the unit of analysis that is grouped by users, instead of 'session' that are mainly used in the transaction-log studies. Clustered Query is a meaningful unit that shows the duration of attention toward a topic and the steps of search. Log Catcher installed in the participants' personal computers and collect logs of the natural web searching behavior. The log data provides quantitative information to the researchers such as duration, a number of 'Clustered Query' and a number of queries in each 'Clustered Query' on a day. It also helps participants answer the questionnaire and helps researchers obtain qualitative information.

III. METHODOLOGY: QUERY CLUSTER

In this section, we describe the methodology, query cluster, to collect user contexts in order to understand the user's intentions in web searching. The first part of this section presents the three phases of the research model and the terms frequently used in this paper. Each phase contains a description of tools that we have developed for the research. For the next part, the pilot and main study that we carried out to validate the methodology is introduced.

Our research model consists of three phases: (1) the Setup Phase, (2) Experiment Phase, and (3) Revision Phase. For each phase, we developed tools to acquire user contexts and intents of web searching behavior (see Figure 1). We borrowed the theories of Marchionini and Jones and Brown

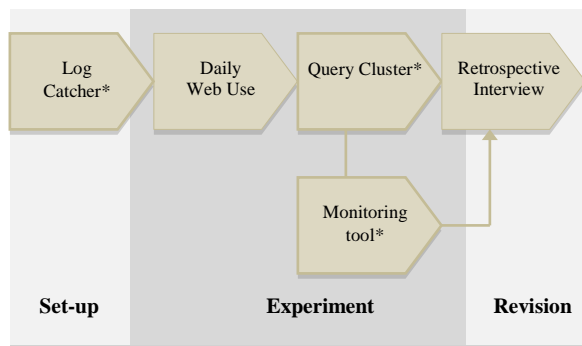


Figure 1. Research Model: 3 Phases

and categorized user contexts into 6 factors, which are information seeker, setting/physical context, task, search system, domain knowledge, and outcomes [16][17]. User data of 6 factors are mainly collected in the experiment phase, and other phases support to fill in the missing information.

For the Set-up phase, participants are given a package consisting of a questionnaire, a log catcher software, and research guidance video. After 1-2 days of stabilizing and learning period, the experiment phase starts. Participants are expected to use their computer daily as usual for two weeks and to access the research website for the assignments, which are clustering the queries of the previous day and answering questions. Researchers monitor participants' daily assignments and note the clustered queries to ask the context. In the revision phase, researchers carry out in-depth-interviews of participants about the clustered-queries and obtain rich user context.

Frequently used terms are defined as follows:

Queries are the keywords that were typed by the user during the process of web portal search activities. In this paper, we collected queries from 7 popular search engines (Google, Yahoo, Naver, Daum, Nate, Paran, and YouTube) in Korea. We focused on the queries rather than the general web activity logs because a series of queries represent user's

information problems and reflect user's wide-ranging information needs [18].

General web activity logs are the website logs that users visited except for the search engine queries.

Cluster is the action of grouping a number of queries that are made under the same objective or intent.

Clustered-queries are the cluster that participants made with their queries collected on the previous day. Each cluster has one or more queries and is titled with subject-representative words. Kelly and Kellar et al. applied similar methods to make participants categorize their own web logs one by one [14][15].

Assignment consists of *Cluster*-ing and answering to the questionnaire about the *Clustered-queries*. A participant is supposed to complete a daily assignment about previous day's web activities log.

A. Set-up Phase

During the set-up phase, researchers understand the basic information about users' web behavior, and participants install the provided tool to their computing environments and learn the tasks for the experiment. An experiment package is provided to the participants, which consists of a guidance video on the tasks and the flows of the experiment, the log catcher tool, and the entry questionnaire. Researchers communicate with participants to inform whether the log data are collected well and to train them how to cluster.

1) Entry Questionnaire

Participants are expected to fill out entry questionnaires at the start of the research. The questionnaire contains 10 items relating to demographic, Internet usage, information ground on the web, and web searching behavior. The questionnaire result provides an understanding of the web search behavior of participants, and some items such as computer type, the period of web usage, the objectives of web search can be used to screen participants.

2) Log Catcher Installation

A client-side log-collecting tool, *Log Catcher* (Log Catcher was written in C# and used windows process hooking mechanism), is published via the research website in a packaged wizard format. As the objective of this research is to understand users' web activities in natural settings, participants are encouraged to install Log Catcher on their personal computer. Log Catcher is designed to collect web activity logs and the condition of collecting web activity logs is informed during the installation process. The following information is collected through Log Catcher:

- Access Time: the time that the participant visited the web site
- Page Title: head title of web page

TABLE I. FACTORS AFFECTING INFORMATION SEEKING

Marchionini (1997) / Jones & Brown (2004)	Research Model		
	Set-up	Experiment	Revision
Information Seeker	-	Motivation	
Setting/Physical Context	Search assistance	Trigger (Physical-context)	Physical-context
Task	-	Clustered-queries	
Search System	Information Source	Search engine Search assistance	
Domain knowledge	-	-	Interview
Outcomes	-	Satisfaction	

TABLE II. INFORMATION COLLECTED BY LOG CATCHER

Access Time	Page Title	URL	Query	Search Engine	Search service	Browser ID	Web Browser
2010-07-22 12:03:27	Sports Today	http://stoo.asiae.co.kr				1114682	Internet Explorer
2010-07-22 12:04:55	Web Hard™	http://www.webhard.co.kr				66218	Internet Explorer
2010-07-22 12:06:09	Naver Search	http://search.naver.com	Sports News	Naver	Web Search	131428	Internet Explorer
2010-07-22 12:06:44	Naver	http://www.naver.com				131428	Internet Explorer
2010-07-22 12:06:45	Naver Search	http://search.naver.com	Free Hi-Pass	Naver	Web Search	131428	Internet Explorer

- URL: URL of visited web page
- Query: typed keywords on the search engine
- Search Engine: search engine that the participants accessed to query
- Search Service: specific search service provided by the search engine (e.g., web search, image search, news search)
- IP address: participant’s IP address
- Web Browser: Web browser’s process ID to distinguish the different web browser windows and tabs
- Web Browser Name: Web browser name that the participant is using (e.g., Microsoft Internet Explorer, Mozilla Firefox, Google Chrome)

Participants should input their activation code provided with the installation package to finish the installation process. The activation code enables researchers to track the participant’s status such as whether the participant was successful in installing the Log Catcher and whether the log data was sent to the server. After installing, Log Catcher is launched automatically when the participant turns on his/her computer and the data collected are sent to the research database every 5 minutes.

Log Catcher stores the participant’s log data until the data is sent to the database to prevent data loss in cases of network failures or unexpected system shutdowns.

B. Experiment Phase

When Log Catcher is stabilized on the participants’ computer, the Experiment Phase follows. Participants are guided to explore the web daily in the same way as they did previously. The daily web activities are collected in the database, and the participant visits the research website to do their assignment. The assignment is consisted of two parts: Cluster and Questionnaires for the Clustered-queries. Participants view their own web activity logs on the research website and cluster the queries in groups according to the rules that they learned. After the clustering finishes, the participant answers to the questionnaire corresponded to the clustered-queries.

1) Definition of ‘day’

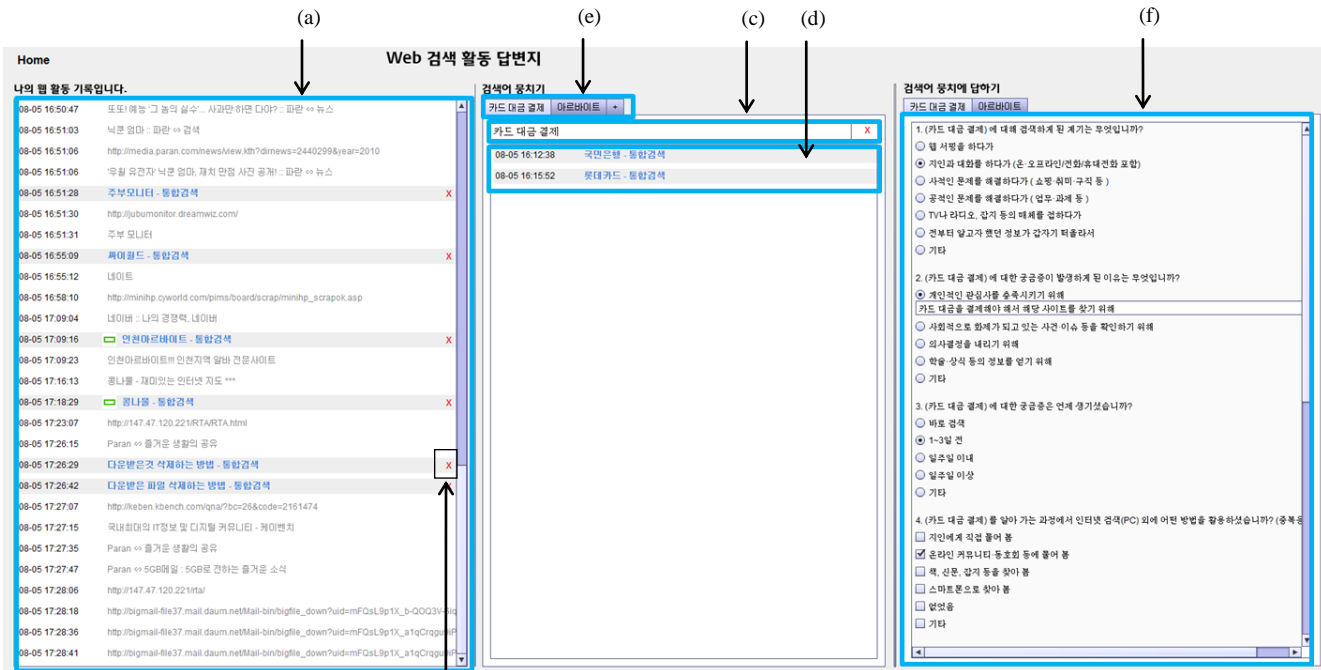
We defined that a ‘day in web usage’ is between 5:00 a.m. and 5:00 a.m. the next day. Due to the number of participants surfing the web until dawn just before they go to work or school, 12:00 am to 12:00 am collection may bring problems for participants in clustering their queries. r their own queries.

2) Query Cluster

Participants are instructed to access the research website and to cluster their own queries that were collected on the previous day. For this, we built a specialized web-log clustering and reporting tool, *Query Cluster*, to help participants to cluster and answer the questionnaire easily. As the design of the Query Cluster is similar to the web card-sorting tool, participants reported that they have no difficulties in using the tool and the tool helped them understand how to cluster their own queries.

When logging in to the Query Cluster with his/her activation code, a participant can view the experiment date page and click the previous date for the assignment. The date is activated if a participant missed the assignment and deactivated if he/she completed the assignment or no logs have yet been collected.

- (a) My web activity logs: The web activity logs of the selected date are shown on the left column. It contains the logs of all websites visited and shows the page title, query texts that user typed, search service, search engines and domains. The web activity logs are listed chronically, and we intentionally left time blanks for any web activity logs that are not collected for an hour. Time blanks and general web activity logs other than the search engine queries are provided for the participants to help them to recollect the reasons why they typed the queries and to cluster the queries. Query logs are lightly shadowed to distinguish from general web activity logs, and the search engine logo is displayed with query logs.



(b) Figure 2. Interface of Query Cluster

- (b) Query Delete button: Participants can remove their own queries from the list in case they do not want to cluster the queries. We added this function after the pilot study as it was observed that several queries were too private to share with researchers or sometimes misspelled. For every deleted query, the reason why it needs to be deleted was specified in order to prevent abusive usage.
- (c) Title of Clustered-queries: The Clustered-queries is titled with subject-representative words that participants typed. (e) represents the titles of Clustered-queries that a participant made.
- (d) Clustered-queries: Participants drag and drop the queries from the left column to the right column to make a cluster. Queries that are moved to the right column are disabled on the left column.
- (e) Titles of Clustered-queries: Participants can make a new cluster by clicking ‘+’ button. A blank title bar and box for clustering appears.
- (f) Questionnaire: Each cluster has a questionnaire and participants should fill out the questionnaires to finish the daily assignment. The details are explained in the ‘contextual questionnaire’ section.

3) Contextual Questionnaire

When participants click the Next button after finishing the Cluster-ing, the contextual questionnaire is displayed. We used five variables to obtain the user data of factors affecting the information seeking:

- **Motivation** is the goal that participants want to reach by resolving the recognized information gap. The multiple choices consists of 5 items that are

derived from the past information behavior studies, which are sense making/reducing uncertainty, decision making, problem solving, fact knowing for personal reason and others. A text area appears when a participant selects an answer for the details of motivation.

- **Trigger** is the internal or external cue for a participant to conduct the type of query. 6 multiple choices are provided for selection, which are during web surfing, communicating with others, working, media consuming, personal affairs, and others. We also asked participants when their motivation arised for the first time.
- **Physical Context** is the user’s environmental context which includes where, with whom, doing what. Although we asked participants to answer the question about physical contexts, we dropped the question as the other questions or interviews could provide the physical contexts.
- **Search assistance features** are the additional functions that a participant uses during the process of search. Choices include auto-completion, related keywords, real-time issues, spell-check, and nothing.
- **Satisfaction** consists of three components: how much the participant gets satisfied, whether a participant reaches the goal through search behavior, and whether he/she would try additional searches later.

C. Revision Phase

In the revision phase, researchers review the collected data and conduct in-depth-interviews to fill the missing clues for the user context.

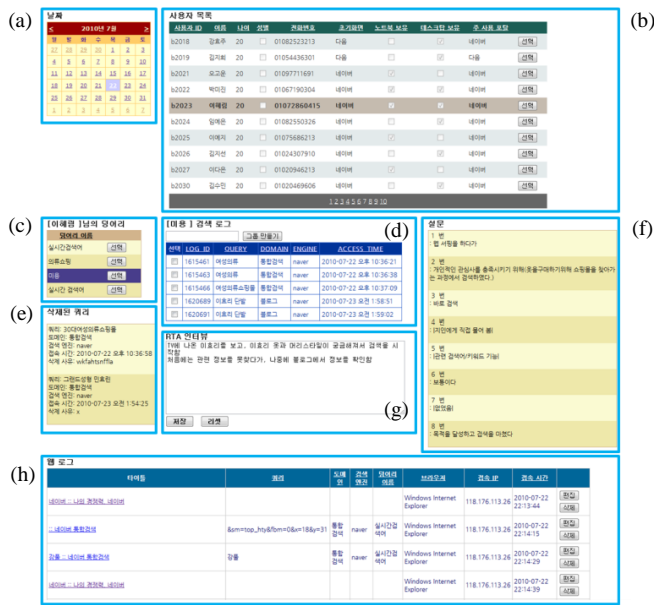


Figure 3. Interface of Monitoring Tool

1) Monitoring Tool

We developed a monitoring tool for researchers. The monitoring tool enables researchers to check the status of research, to extract unusual queries/answers at a glance and to interview with well-formatted data. The interface of this tool is shown in Figure 3.

- (a) Calendar: Researchers can select the date to check.
- (b) List of participants: Participants whose data were collected on the selected date are listed. The list contains id, demographic information, and background information of participants. (e.g., name, age, gender, phone number, search engine preference).
- (c) Titles of Clustered-queries
- (d) Clustered-queries: Clustered-queries are listed when a researcher selects a title of clustered-queries. The information contains query text, search engine, searched domain, and access date and time. Researchers can create or modify clusters if the participant asks to change.
- (e) Deleted queries: The deleted queries and reasons appear if the participant removes the query from the logs.
- (f) Questionnaire answers about the cluster
- (g) Memo box: Researchers can leave notes about the findings during the interview.
- (h) Web activities log: Contains both queries and general web activity logs.

2) Exit Interview

A retrospective interview technique is applied for the exit interview. Although the technique is not considered to

be appropriate to collect accurate and objective data, it is useful to build a history of event or exploratory experience by making participants recall their aspects of past experiences [19].

Researchers carry out the exit interview via telephone once a week and twice during the 2-week research period. Before the interview, researchers reviewed the Clustered-queries and Contextual Questionnaire of users. Participants are asked to tell a story about the situation and motivation of Clustered-queries and the relations among the queries in a cluster.

Participants' domain knowledge about the clustered-queries also can be asked through the exit interview: whether a participant is accustomed to the topic, which information grounds he/she relies on for seeking information about the topic, or how frequently a participant searches for the related topic.

IV. RESEARCH PROCESS

The entire research process is designed to modify problems and to prove the validity of the methodology. Several changes have been made through the pilot study, and tools and survey questions were reviewed. After refining the methodology, we carried out the main study with a large number of participants and found concerning points when applying the methodology.

A. Pilot Study

Among 8 participants recruited for the pilot study, 4 were females and 4 were males and all participants were between the ages of 20-40. We considered the participants' job, where 4 were undergraduates and 4 were paid workers. During the 6-day pilot study, most participants set up the log catcher tool in their personal computer because of the security issues in collecting logs at their workplace computers. The participants had selectively sent their web search logs by turning on and off the log catcher tool. Participants clustered their log histories on the web, and answered survey questions about each log clusters on a daily basis. At the end of the pilot study, the researcher interviewed the participants via mobile phone for 30 minutes or more to acquire user contexts in web searching activities.

B. Problems

We found several problems through the pilot study and modified the research process and details to enhance the participants' engagement and to acquire valid user data from the study.

1) Selective Report of Log Histories

In the pilot study, participants selectively reported their log histories by clicking the on-off button provided. We asked participants to turn on the log catcher tool at least two hours a day to acquire equivalent amount of data from all subjects and to respect their privacy. However, some

participants got confused when turning the log catcher on and failed to report their search histories. Other participants intentionally hid their search histories and invented data for the report. The log catcher tool was modified to collect the whole log histories during web use. In the main study, we recruited participants who accepted this condition and added a function to delete queries on the tool.

2) *Erros in Clustering*

Throughout the pilot study, we noticed that participants made mistakes when clustering their own queries at times. Some participants, for example, classified queries into different clusters although the queries were made with the same motivation. Other examples are inappropriate titles, for example, a participant titled clustered-queries as 'day 1'. We trained those participants to cluster queries based on a motivation, and to title clustered-queries to represent the subject of the cluster.

Most errors were found in the early stages of the experiment, as participants are not accustomed to the experiment and the clustering rule at first. Therefore, the data of first 1~2 days should be reviewed carefully and researchers should communicate with participants to understand how to cluster and title. However several cases were reported during the interviews and it was also required to provide a cluster-modification function on the administration tool. In the main study, researchers combine or separate queries during the interview.

3) *Surveys about Clustered-queries*

The initial version of the contextual questionnaire contains several open-ended questions to ask participants to answer the physical contexts or search reasons. However, we found that the participants were not willingly answering the open-ended question of each clustered queries. We decided to drop the burdensome questions as the exit interview study and other questions can cover them. We also found that our participants usually search at home, the physical contexts are not considered as crucial feature in this study.

C. *Main Study*

After revising the methodology, we carried out the main study with a large number of participants. 100 participants were recruited for the main study and the demographic ratio of participants were similar to the Korean demographic data except for the geographic; 25 male undergraduates and 25 female undergraduates in their 20's, and 25 male paid workers and 15 female paid workers-10 housewives in their 30's.

For 14 days of the main study, most participants installed the log catcher tool on their personal computer. Entire log histories of the participants were collected on the server and participants were asked to cluster queries and to answer the surveys to each clustered queries of the previous day. During the main study, participants were interviewed

twice about the context of the clustered queries or correlation between the clustered queries found on another day.

V. CONCLUSION

This paper presents a methodology to understand the user context in web search behavior with Clustered Query as a research unit. The methodology consists of three phases - set-up, experiment, and revision phase-, and the methodology is refined and validated through the pilot and main study. User context is defined as 6 factors and tools are introduced that are developed to obtain user data in each phase.

The contributions of the proposed methodology are usability and user-oriented approach. Participants highly engage in the experiment phase by clustering their own queries, and provide meaningful clusters that cannot be captured through previous log analysis studies. Our methodology improved the previous quantitative and qualitative approaches by collecting quantitative data of users' web activities logs and qualitative data of questionnaires and interviews. The self clustered-queries deliver valuable data to understand the user intents and the task session.

For the next step, we will analyze the data obtained through the pilot and main study focusing on the categorization of user intent and its effect on search behavior.

REFERENCES

- [1] The Mainstreaming of Online Life. http://www.pewinternet.org/~media/Files/Reports/2005/Internet_Statut_2005.pdf.pdf, 01.07.2011
- [2] <http://www.google.com>, 01.07.2011
- [3] <http://www.naver.com>, 01.07.2011
- [4] B. Jansen and A. Spink, "How are we searching the World Wide Web? A comparison of nine search engine transaction logs," *Information Processing & Management*, vol. 42, Jan. 2006, pp. 248-263, doi:10.1016/j.ipm.2004.10.007.
- [5] C. Silverstein, H. Marais, M. Henzinger, and M. Moricz, "Analysis of a very large web search engine query log," *ACM SIGIR Forum*, vol. 33, Fall 1999, pp. 6-12, doi:10.1145/331403.331405.
- [6] B. Jansen, A. Spink, and J. Pedersen, "A temporal comparison of AltaVista Web searching," *Journal of the American Society for Information Science and Technology*, vol. 56, Feb. 2005, pp. 559-570, doi:10.1002/asi.20145.
- [7] N. Ross and D. Wolfram, "End user searching on the Internet: An analysis of term pair topics submitted to the Excite search engine," *Journal of the American Society for Information Science*, vol. 51, Jun. 2000, pp. 949-958, doi: 10.1002/1097-4571(2000)51:10<949::AID-ASIT70>3.0.CO;2-5.
- [8] X. Shi and C. Yang, "Mining related queries from Web search engine query logs using an improved association rule mining model," *Journal of the American Society for Information Science and Technology*, vol. 58, Aug. 2007, pp. 1871-1883, doi: 10.1002/asi.20632.
- [9] D. Rose and D. Levinson, "Understanding user goals in web search," *Proceedings of the 13th international conference on World Wide Web (WWW 04)*, ACM, 2004, pp. 13-19, doi:10.1145/988672.988675.
- [10] A. Broder, "A taxonomy of web search," *ACM SIGIR Forum*, vol. 36, Fall 2002, pp. 3-10, doi:10.1145/792550.792552.

- [11] B. Jansen, A. Spink, and T. Saracevic, "Real life, real users, and real needs: a study and analysis of user queries on the web," *Information Processing & Management*, vol. 36, Mar. 2000, pp. 207–227, doi:10.1016/S0306-4573(99)00056-4.
- [12] C. Choo, B. Detlor, and D. Turnbull, "Information Seeking on the Web: An Integrated Model of Browsing and Searching," *Proceedings of the ASIS Annual Meeting, Information Today*, vol. 36, Oct. 1999, pp. 3–16.
- [13] A. Sellen, R. Murphy, and K. Shaw, "How knowledge workers use the web," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems(CHI 02)*, ACM, 2002, pp. 227-234, doi:10.1145/503376.503418.
- [14] D. Kelly, "Measuring online information seeking context, Part 1: Background and method," *Journal of the American Society for Information Science and Technology*, vol. 57, Sep. 2006, pp. 1729–1739, doi:10.1002/asi.20483.
- [15] M. Kellar, C. Watters, and M. Shepherd, "A field study characterizing Web-based information-seeking tasks," *Journal of the American Society for Information Science and Technology*, vol. 58, Mar. 2007, pp. 999–1018, doi:10.1002/asi.20590.
- [16] G. Marchionini, "Information Seeking in Electronic Environments," Cambridge University Press, 1995, pp. 33-38.
- [17] G. Jones and P. Brown, "The role of context in information retrieval," *Proceedings of the ACM SIGIR 2004 Workshop on Information Retrieval in Context*, ACM, 2004, pp. 20-22.
- [18] S. Rieh and H. Xie, "Analysis of multiple query reformulations on the web: The interactive information retrieval context," *Information Processing & Management*, vol. 42, May 2006, pp. 751–768, doi:10.1016/j.ipm.2005.05.005.
- [19] B. Montgomery and S. Duck, eds. "Studying Interpersonal Interaction. Guilford Publishing", The Guilford Press, 1991, pp. 162-178.