How Do Users Grow Up along with Search Engines?: A Study of Long-term Users' Behavior

Jian Liu, Yiqun Liu, Min Zhang, Shaoping Ma
State Key Laboratory of Intelligent Technology and Systems
Tsinghua National Laboratory for Information Science and Technology
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China
liujian2014@gmail.com, yiqunliu@tsinghua.edu.cn, z-m@tsinghua.edu.cn, msp@tsinghua.edu.cn

ABSTRACT

With a stronger reliance on search engines in our daily life, a large number of studies have investigated user behavior characteristics in Web search. However, previous studies mainly focus on large-scale query log data and analyze temporal changes based on all users without differentiating different user groups; few have really traced a fixed and long-term group of users and have distinguished the behavior of long-term users from ordinary users to analyze long-term temporal changes unbiasedly. In this paper we look into the interaction logs of these two user groups to analyze differences between these two user groups and to better understand how users grow up along with Web search engines. Statistical and experimental results show that there exist temporal changes of both user groups. There are also significant differences between these two user groups in the frequency of interaction, complexity of search tasks, and query formulation conventions. The findings have implications for how Web search engines should better support users' information seeking process by tackling complex search tasks and complicated query formulations.

Categories and Subject Descriptors

H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval

Keywords

User behavior analysis, Web search engine, Interactive degree, Search task complexity, Query formulation conventions

1. INTRODUCTION

The rapid and continuous growth of the World Wide Web has made the Internet an indispensable part of our daily lives. According to *Internet World Stats*, on June 30th 2012, Internet users from all over the world reached approximately 2,405 million, i.e., 34.3% of the world population [1]. With the explosively increasing of Internet users, the amount of data on the Web is also expanding rapidly. That is why people must resort to Web search engines to locate information quickly and precisely. As the Internet users have different skill levels (i.e. newbies and advanced users) and are performing different search tasks (e.g., simple and complex tasks), many researchers in the area of Information Retrieval have tried to identify the average characteristics of user interactions and temporal evolutions in user behavior. So that includes improving the performances of search engines by providing useful results, enhancing users' search experiences, and so on.

However, most studies mainly focus on large-scale query log data, and identify the average characteristics of user interactions without differentiating Internet newbies from long-term users. Studies on different user groups have already shown that search engine users are a heterogeneous crowd and need to be dealt with differently. There are significant differences in querying, result-clicking, and other search behaviors in the crowd [2]. Therefore, many previous works have limitations in studying long-term temporal changes of search engine users' behavior, and their results cannot reflect the real characteristics of long-term users' behavior if they didn't trace a fixed group of users who had lasted for several years. In this paper, we try to differentiate long-term users from ordinary users, and trace the evolution of long-term user interactions with Web search engines.

Generally speaking, an information seeker becomes more skillful in performing search tasks if he uses Web search engines longer than other users. In order to unbiasedly investigate such temporal changes of search engine users' behavior, we should trace a fixed group of users who have used the same Web search engine for several years, and then study the behavior of these users to yield important insights into the evolution of users' Web search activities. Based on this idea, we collect large-scale and long-term query logs from a popular commercial Web search engine, and then differentiate query logs of long-term users (identified by long-lasting users' unique identification) from ordinary users' query logs, and statistically and experimentally investigate their differences in the frequency of interaction, complexity of search tasks, and query formulation conventions.

To describe the evolution of how users interact with search engines, we will answer three research questions as follows:

(1)How does the interaction process of search users evolve during a time period of several years?

(2)How does the complexity of search tasks of long-term users change over time?

(3)How do query formulation conventions of long-term users change over time?

Through statistical and experimental studies and analysis, we will give potential answers to these questions. All the long-term temporal changes reflect the evolution of users' Web search activities, and their differences have possible implications for Web search engines to better support users' information seeking behaviors.

The rest of this paper is structured as follows: In Section 2 we review related work in query log analysis. The information of query log data sets and detail of identifying long-term users are presented in Section 3. The fourth section focuses on the average characteristics of user interactions and temporal changes of user behavior. We discuss the implications of this research in Section 5 and conclude this work in Section 6.

2. RELATED WORK

Since late 1990s many researchers have investigated how users use Web search engines to satisfy their information needs. By analyzing query logs recorded by search engines, many characteristics of users' searching activities have been discovered. Silverstein et al. [3] analyzed individual queries, query duplication, and query sessions based on an AltaVista Search Engine query log. They reported that for 85% of the queries only the first result screen was viewed, and that 77% of the sessions contained only 1 query. Jansen et al. [4] conducted an in-depth analysis of query logs with the Excite search engine, and reported that user sessions were short and only a small number of terms were used in a query. According to their analysis, approximately 54% of Web users viewed only one page of results, and users on average viewed about 2 to 3 documents per query, and over half the sessions were less than 15 minutes [5]. Guy et al. provided an extensive analysis of how and why people search is used within the organization *Faces* based on a large-scale query log including more than a million queries [6].

However, the above characteristics only describe general statistical features of users' behavior, and can't describe how users interact with search engines in different user groups. Following this direction, many different researches have been conducted based on different criteria. Spink et al. [7] examined users' behavior depending on user's geography location information. A comparative study showed that there were significant differences in Web search behavior of the United States and European users. Chau et al. [8] investigated differences between general-purpose search engine, such as Excite and AltaVista, and in-site search engine, which allows users to search for pages only within a particular Web site. Their results showed that there were considerably different characteristics such as search topics and query terms used.

Besides analysis of the static characteristics, researchers also want to identify the temporal evolution of users' Web search activities. Jansen et al. conducted an analysis to study long-term temporal changes of European Web search behavior, and reported that there were some declines in query length, notable percentage decreases in search for sexual and pornographic contents [9]. They also performed an in-depth analysis of large-scale and long-term query logs, and their results showed that there were increases in session and query length, 70% of session durations were less than 5 minutes, and Web searchers' information needs became broad with the most frequent terms accounting for less than 1% of total term usage [10]. Beitzel et al. analyzed topically categorized queries on a daily, weekly, and monthly basis to identify topical trends, and concluded that certain categories of queries trend differently over varying periods [11]. Pang et al. examined the temporal changes of question-like query formulation, and showed that Web users were more likely to use natural questions to express their search intent as time went by [12].

Other studies concerning users' behavior have attempted to focus on different user groups' behavior. It is believed that search engine users are a heterogeneous crowd and we should sort out different categories of users before we analyze their behaviors. Hölscher et al. [2] grouped users into two categories: *experts* and *novices*, and examined their searching behaviors. They found that *experts* relied more on using their own keyword-based queries to search, but *novices* were more likely to use other users' tags to search. White et al. reported advanced search engine users' behavior based on a log-based study, and concluded that there were significant differences between advanced users and other users [15]. Kang et al. [13] compared search behavior of experts and novices, and concluded that experts relied more on their own domain knowledge to generate search queries.

All the above studies either focus on identifying long-term temporal changes of users' behavior without distinguishing long-term users from ordinary users, or only sort out different types of users for other usages. To unbiasedly investigate the long-term temporal changes of search engine users' behavior, we trace a fixed group of users who have used the same search engine for over three years, analyze their behaviors, and compare with ordinary users to get insightful conclusions.

3. DATASET AND CHARACTERISTICS

3.1 Dataset

With the help of a popular commercial Web search engine company in China, we obtain three query log data sets from a fixed time of year– September 1st to 14th– in three consecutive years from 2009 to 2011. In total, there are about 760,794,583 records in the three query log files. The information contained in each record includes a timestamp of each page viewed, an anonymous cookie assigned by the search engine system to identify a unique user, a query submitted by the given user, and the URL clicked by him. To distinguish long-term users from ordinary users, we define these two user types as follows:

- **Long-term user**: a user is a *long-term user* if he appears at least once in each period of time, namely the user must use the same search engine system at least three years.
- **Ordinary user**: all users except long-term users.

We statistically investigate query logs from three different periods of time, and the numbers of unique users are 60,443,119; 92,087,410; and 129,534,711, respectively. Among them, 18,521 users (0.0306%, 0.0201%, and 0.0143%, respectively) are confirmed as long-term users.

The privacy of users is maintained throughout the whole process of the study: No personal information such as IP address or locality is included. Users are assigned a unique anonymous identifier that could not be traced back to the users. Meanwhile, all the characteristics of user behavior are based on multiple users without attempting to identify a particular users' behavior.

To find out more about temporal changes of long-term users' behavior and user behavior differences between long-term users and ordinary users, we focus on several search behavior features including queries, result-clicks, and search tasks.

3.2 Search Behavior Features

Typically, during a Web search session, a user takes several actions sequentially including issuing a query, viewing result pages, clicking on URLs, viewing documents, returning to the search engine for query reformulation and so on. Generally speaking, all these actions are based on a single information need: a search task. Queries, result-clicks, and search tasks are the fundamental features to describe users' Web searching activities. In the following paragraphs, we describe the characteristics of these features in more details.

Queries are submitted to Web search engines. They represent users' search intent and are treated as important resources to observe user interactions. We examine statistics of queries such as the length of an individual query, the information need of an individual query, and query distributions across different user groups. Especially, we focus on queries which describe information needs by natural-language questions. We choose the selection strategy adopted by Pang et al. to identify question-like queries [12].
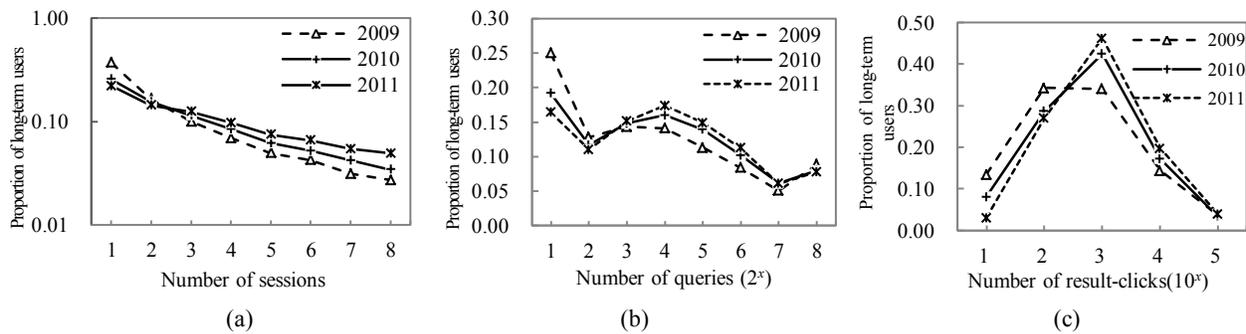
Figure 1: The proportion of long-term users over years, with respect to the number of (a) sessions, (b) queries, (c) result-clicks.

Result-clicks have potential meanings for studying user behavior and are always considered as implicit feedback from users. Because of the advantage of low cost and large quantities, implicit feedback has been getting more attention from researchers. In this paper, we inspect result-clicks from several dimensions such as query-level and search-task-level.

In order to identify a **search task**, namely, an information need, we adopt the definition of Jansen et al. for a session: A series of interactions by the user toward addressing a single information need, and choose a session inactivity timeout of 30 minutes to demarcate two sessions of the same user [14, 15]. In the rest of this paper, we use session metrics to characterize search tasks. Several characteristics of sessions such as the average duration of users' sessions, the average number of queries and of result-clicks in a session are discussed.

All the features described above allow us to investigate long-term temporal changes of users' Web searching activities, as well as user behavior differences between long-term users and ordinary users. Hence they allow us to better understand how users grow up along with Web search engines.

4. HOW LONG-TERM USERS PERFORM SEARCH TASKS

In this section, we describe how users grow up along with search engines through quantitative analysis of query logs, as well as experimental results. The analysis is divided into three parts: (1) Analysis of interactive degree between users and Web search engines: How frequently do users interact with search engine system. (2) Analysis of the complexity of search tasks. (3) Analysis of query formulation conventions: Are there any differences in query formulation conventions between long-term users and ordinary users. That is to say, we mainly focus on temporal changes of all the three parts and differences between long-term users and ordinary users.

4.1 Evolution of Interactive Degree

The number of individuals relying on search engines to solve their specific problems has been increasing. Firstly, we are interested in investigating whether users become more interactive with Web search engines as time goes on. Thus, we examine how frequently users interact with Web search engines, and compare differences in interactive degree between long-term users and ordinary users. When a user interacts with a search engine system, the whole process involves sessions, queries, and result-clicks. So we measure the *interactive degree* from the three aspects: (a) Proportion of users vs. number of sessions, (b) Proportion of users vs. number of queries, and (c) Proportion of users vs. number of result-clicks in a fixed period of time. A higher *interactive degree* means that users interact with a search engine system more frequently.

Figure 1 shows the temporal changes of *interactive degree* of long-term users in three dimensions from 2009 to 2011. Figure 1(a) describes the proportion of long-term users as a function of the number of sessions in a fixed period of time. As depicted, more than 90% of long-term users only have at most three sessions. Moreover, the proportion of long-term users with one or two search sessions has declined, while the proportion of long-term users with more than two sessions has increased from 2009 to 2011. That is to say, more percentage of users has issued much more sessions in the fixed period of time from 2009 to 2011. Figure 1(b) describes the proportion from the dimension of queries; it shows that the proportion of long-term users has increased when the number of queries is equal or greater than 8 (2^3) from 2009 to 2011. In the feature of result-clicks shown in Figure 1(c), we find that when the number of result-clicks is more than 100 (10^2), the percentage of long-term users has increased over years. All the temporal changes shown in Figure 1 indicate that long-term users are tending to have more search sessions, queries and result-clicks as time goes on, that is to say, users interact with Web search engine systems more and more frequently. To validate the significance of differences, we performed a Student's two tailed *t*-test (P<0.05) and all differences are significant.

To learn more about differences between long-term users and ordinary users in the interactive degree with Web search engine systems, we investigate their differences on average from the perspectives of the above three features in Figure 2, and "Proportion of users" on the y-axis means that for each user category, summing over all data points, it will be 1. Although we omit the long tail of each graph, it is very small and doesn't affect the conclusions. In Figure 2(a), to render the differences more clearly, we adopt the log scale y-axis. The proportion of users as a function of the number of sessions shown in Figure 2(a) presents that long-term users tend to have more sessions in a fixed period of time. As we can see in Figure 2(b), with the increasing number of queries, the proportion of ordinary users is declining sharply, while the proportion of long-term users remains relatively high. Similar phenomena occur in result-clicks. Shown in Figure 2(c), although about 10% (and more 40%) of long-term users whose result-clicks are less than 10 (and 10^2) in the fixed period of time, they performed more result-clicks compared with ordinary users from the overall distribution of result-clicks. We also performed a paired two tailed *t*-test, the differences of all the features are significant based on the fact that the absolute value of the *t*-test statistic (*t-stat*) is larger than the observed absolute value of the *t*-test statistic (|*t-stat*|>1.96 namely P<0.05).

All the quantitative analysis above show that long-term users have become more interactive as time goes on, and that have a higher interactive degree when they interact with Web search engines compared with ordinary users.
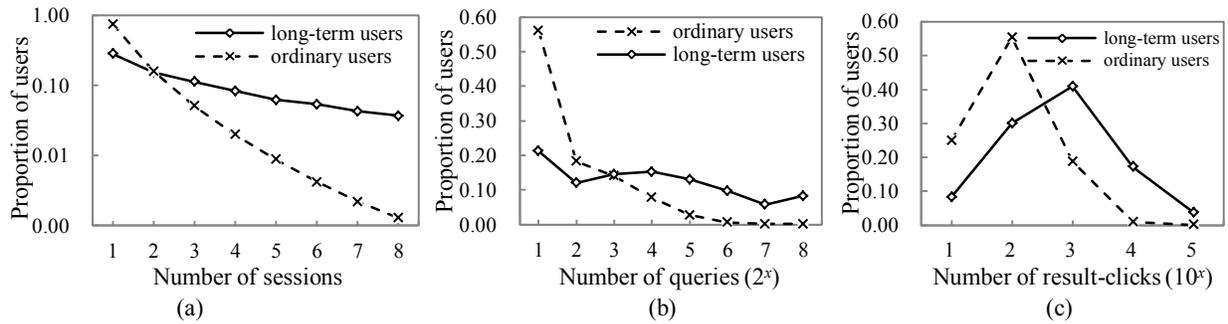
Figure 2: The proportion of users with respect to (a) Number of sessions, (b) Number of queries,(c) Number of result-clicks.

Table 1. Number of queries per session on average for long-term users and ordinary users.

Year	Long-term	Ordinary
2009	7.37	2.20
2010	4.04	1.95
2011	3.47	1.83

Table 2. Average session duration per session (in sec.) for long-term users and ordinary users.

Year	Long-term	Ordinary
2009	2,595	838
2010	1,451	740
2011	1,281	694

Table 3. Percentage of sessions which include more than one result-click over different years.

Year	Long-term	Ordinary
2009	70.1%	60.0%
2010	71.9%	57.4%
2011	70.5%	56.2%

4.2 Evolution of Complexity of Search Tasks

Many studies have been conducted to group queries into different search tasks or sessions. White et al. investigate two extreme user groups: *explorers* and *navigators*, and find that users submit many queries during a search session and that they visit many new domains in complex sense-making tasks, whereas search interaction is highly consistent in well-defined fact-finding tasks [16]. Aula et al. statistically confirm that when users have difficulties in a search task, they spend more time on the search result page [17]. Based on these studies, we can infer that a more complex search task has more queries in a session, longer session duration, more result-clicks, etc. Therefore, we look into the complexity of search tasks from: the number of unique queries in sessions, the session duration, and the distribution of result-clicks in sessions.

Table 1 shows the average number of unique queries per session. Obviously, the number of unique queries per session used by long-term users is consistently more than ordinary users from 2009 to 2011. For example, the number of unique queries on average is 3.47 compared with 1.83 used by ordinary users in 2011, and there is an even larger difference of 7.37 compared with 2.20 in 2009. All differences are significant (*p*-value is less than 0.05 with a Student's two tailed *t*-test). Furthermore, our results also show that the number of unique queries per session used by either long-term users or ordinary users has declined over years.

Generally speaking, users tend to spend more time on completing a more complex search task. The average session duration time is provided in Table 2. Average time spent on sessions by long-term users is consistently longer than that by ordinary users. In 2009, the time spent by long-term users is 2,595 seconds, compared with only 838 seconds by ordinary users. Furthermore, search engine users are tending to spend less time on search tasks on average as time goes on, and a similar result is provided by previous work [5]. Additionally, the difference between these two user groups is significant based on the value of a Student's *t*-test statistic (|*t-stat*|>1.96 namely P<0.05). This phenomenon can be mainly explained by the fact that the performance of search engines is improved in recent years and that users only need to spend less time on the same level of complexity of search tasks. Moreover, we investigate the percentage of sessions which includes more than one result-click in sessions. Our observations are summarized in

Table 3. The average characteristic shows that search tasks performed by long-term users include more result-clicks than those performed by ordinary users by more than 10%.

All the analyses above show that long-term users tend to perform more complex search tasks when they interact with search engines. To validate our inference, we further conduct an in-depth investigation of search task complexity of these two user groups. Based on the idea of a three-level categorization of search task complexity defined by Bell et al.[18], we categorize the search tasks into two groups: *navigational search task*, which corresponds to *Complexity Level* 1 search task; and *non-navigational search task*, which corresponds to *Complexity Level* 2 and 3. According to inherent relationships between the definition of *Complexity Level* 1, 2, 3 search task and the definition of query-type [18, 19], we can use the query-type (*navigational, informational, transactional*) of all queries in a session to define the two search task types:

- **Navigational search task**: All queries in the search task are *navigational* type queries.
- **Non-navigational search task**: Not all queries in the search task are *navigational* type queries.

Although the definition is naive, it does not affect the evaluation of the complexity of a search task. Meanwhile, *navigational search task* is less complicated than *non-navigational search task* according to the work [18]. Thus, we investigate search task complexity from the distribution of the two search task types in two user groups. We randomly extract 200 sessions from the two user groups, respectively. For all queries in each session, we organize two professional annotators to annotate query-type, and Table 4 shows two user groups' search task types in more details.

Table 4. The distribution of different search task type for long-term users and ordinary users (the number of sessions).

Search task type	Long-term users	Ordinary users
Navigational	35(17.5%)	58(29.0%)
Non-navigational	165(82.5%)	142(71.0%)

In Table 4, we know that long-term users performed less navigational search tasks than ordinary users, of which the percentage is 17.5% compared with 29%. We performed a Student's two tailed
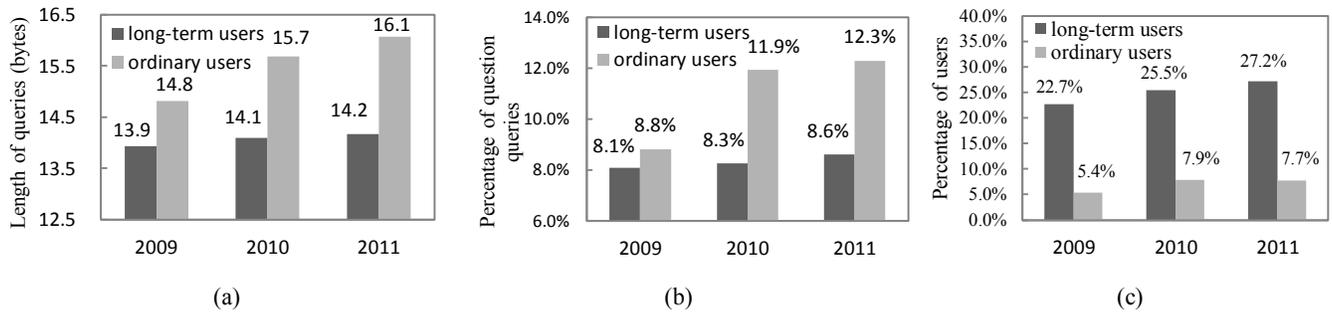
Figure 3: (a) Average length of all unique queries over years in bytes; (b) Percentage of question queries used long-term users and ordinary users; (c) Percentage of users who submitted question queries at least once in a fixed period of time over years.

t-test and the absolute value of the *t*-test statistic (*t-stat*) is 2.72, which is larger than 1.96. So the difference is significant. From the experimental results, we can infer that search tasks performed by long-term users are mostly complex than that by ordinary users. Based on the investigations and experimental results, we can confirm that long-term users rely on search engines to solve more complex problems, not only to navigate to a simple Web page.

4.3 Evolution of Query Formulation Conventions

Query formulation is an essential step in a search session. Many studies have been conducted to investigate how users could formulate effective queries. Aula reports that experienced users formulate more specific and longer queries whereas less experienced users only use fewer and more generic terms to express their information needs [20]. Based on these studies, we focus on two subjects on query formulation: the length of unique queries and the usage of question-like queries.

Question-like queries, as described in the section 3, are natural-language questions posed as queries. We define question-like queries based on the standard definition of questions in Chinese, and identify a question-like query through whether a query contains one of Chinese characters, which are corresponding to interrogative words including *how, what, which, why, where, when, who,* and *whose* in English. Figure 3(b) shows the percentage of question-like queries with respect to two user groups. We find that they account for 8.8% of all unique queries submitted by ordinary users, compared with 8.1% by long-term users in 2009, and the gap between these two groups becomes larger as time goes on. All differences are significant based on the two tailed *t*-test (P<0.05). Meanwhile, the percentage has continuously increased over years in both user groups, which indicates that all users are more likely to use question-like queries to express information needs as time goes on. The trend is consistent with the previous work concerning English-language question-like queries [12].

Furthermore, we investigated the proportion of users who submit question-like queries at least once in a fixed period of time. The results are shown in Figure 3(c), and 22.7% of long-term users submitted question-like queries at least once whereas only 5.4% of ordinary users did that in 2009, and the difference between the two user groups is statistically significant from 2009 to 2011 (P<0.05 with a Student's two tailed *t*-test). As we see in Figure 3(c), more long-term users submit question-like queries over years, which can be explained that Web search engines become more powerful to solve question-like queries by incorporating the Web resources from Web community Q&A, such as Google Answers. However, in Figure 3(b), a smaller percentage of question-like queries are submitted by long-term users, while a larger propor-

tion of long-term users submit question-like queries as shown in Figure 3(c). Intuitively, we consider that these two phenomena are contradictive, but actually, based on our careful investigations into several random users' specific queries, we find that mostly long-term users do have the custom of issuing question-like queries in their search history, but not always. So, it can be explained that, compared with ordinary users, long-term users are more likely to formulate question-like queries in a proper time when their information needs can be suitably expressed by this kind of queries, but not all the time. Additionally, question-like queries, which are natural-language sentences, are longer than keyword-based queries. As discussed above, users issue question-like queries more frequently these years, which well explained the reason why the average length of unique queries increased shown in Figure 3(a).

All of the above mentioned facts show that all users are more likely to use question-like queries to express their information needs over time, and a larger proportion of long-term users use question-like queries, as compared to those of ordinary users. However, long-term users would like to use question-like queries only when they reasonably consider that this kind of queries is more suitable than keyword-based queries to express their information needs. Based on the conclusion that users formulate more question-like queries when faced with difficult search tasks noted in previous work [17], the results well support the conclusion in section 4.2 that long-term users rely on Web search engines to solve more complex problems, not only to navigate to a Web page.

5. DISCUSSION

Our results show that compared with ordinary users, long-term users have different characteristics in Web search behavior:

(1) Long-term users have a higher interactive degree than ordinary users when they use Web search engines, namely, they interact with Web search engines more frequently.

(2) Long-term users rely on Web search engines to solve more complex problems, not only to navigate to a Web page.

(3) Long-term users do have the custom of issuing question-like queries, and compared with ordinary users, much more percentage of them is likely to use question-like queries to express their information needs.

All the findings above may look like independent to each other superficially, but they are inherently consistent with each other in our opinion. With the fact that long-term users are more likely to perform more complex search tasks, as confirmed by our manual investigation, we can explain the consistency of these findings based on the contributions of previous work that when users perform more complex search tasks, they submit more queries and spend much more time, and they tend to formulate more question-like queries when faced with difficult search tasks [16, 17].

Long-term users are the most loyal users of a Web search engine. Thus, how to keep these users is an important issue for Web search engines. A higher interactive degree means more revenue for Web search engines; meanwhile more complex search tasks mean more challenging for them. Specially, it is important to note that more question-like queries used by users pose a big challenge for search engines to understand their search intent. All these findings have implications for Web search engines designers on how to better support users' search process changes, and what Web search engines should do to tackle more and more complex search tasks and complicated query formulation conventions efficiently and effectively.

As with any research, there are limitations that should be recognized. Although an anonymous cookie assigned by a search engine is, to our best knowledge, the only way to identify a unique user in log-based analysis, it actually only can identify a same computer, which may be used by several persons. However, according to the report from CNNIC (China Internet Network Information Center), 90.3% of Internet users access Internet at home through personal computers, not communal computers. So we believe that the impact is small in our research.

6. CONCLUSION

In this paper, we have presented a log-based study of long-term users' behavior, which has uncovered several temporal changes of long-term users' behavior, as well as significant differences in search behavior between long-term users and ordinary users. Long-term users interact with Web search engines more frequently, tend to rely on search engines to solve more complex problems, and have a larger probability to use question-like queries to express information need. As we have suggested, Web search engines should do more, such as better tackling question-like queries, leveraging interactions of long-term users for better performing a more complex search task, to support users' search behavior changes. Future work will include investigating potential benefits for Web search engines' performance, especially the challenge of rare queries and question-like queries, and extracting valid features of long-term users' behavior as indicators of search expertise.

7. ACKNOWLEDGMENTS

8. REFERENCES

[1] Internet World Stats. 2012. World Internet Users and Population Stats. *http://www.internetworldstats.com/stats.htm.*

[2] Hölscher, C., and Strube, G. 2000. Web search behavior of Internet experts and newbies. *Computer networks, 33*(1), 337-346.

[3] Silverstein, C., Marais, H., Henzinger, M., and Moricz, M. 1999. Analysis of a very large web search engine query log. In *ACM SIGIR Forum* (Vol. 33, No. 1, pp. 6-12). ACM.

[4] Jansen, B. J., Spink, A., and Saracevic, T. 2000. Real life, real users, and real needs: a study and analysis of user queries on the web. *Information processing & management, 36*(2), 207-227.

[5] Jansen, B. J., and Spink, A. 2003. An Analysis of Web Documents Retrieved and Viewed. In *International Conference on Internet Computing* (pp. 65-69).

[6] Guy, I., Ur, S., Ronen, I., Weber, S., and Oral, T. 2012. Best faces forward: a large-scale study of people search in the enterprise. In *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems*.1775-1784. ACM.

[7] Spink, A., Ozmutlu, S., Ozmutlu, H. C., and Jansen, B. J. 2002. US versus European Web searching trends. In *ACM Sigir Forum* (Vol. 36, No. 2, pp. 32-38). ACM.

[8] Chau, M., Fang, X., and Liu Sheng, O. R. 2005. Analysis of the query logs of a web site search engine. *Journal of the American Society for Information Science and Technology*, *56*(13), 1363-1376.

[9] Jansen, B. J., and Spink, A. 2005. An analysis of web searching by European AlltheWeb.com users. *Information Processing & Management*, 41(2), 361-381.

[10] Jansen, B. J., Spink, A., and Pedersen, J. 2005. A temporal comparison of AltaVista Web searching. *Journal of the American Society for Information Science and Technology*, 56(6), 559-570.

[11] Beitzel, S. M., Jensen, E. C., Chowdhury, A., Frieder, O., and Grossman, D. 2007. Temporal analysis of a very large topically categorized web query log. *Journal of the American Society for Information Science and Technology*, *58*(2), 166-178.

[12] Pang, B., and Kumar, R. 2011. Search in the Lost Sense of "Query": Question Formulation in Web Search Queries and its Temporal Changes. In *ACL (Short Papers)* (pp. 135-140).

[13] Kang, R., and Fu, W. T. 2010. Exploratory information search by domain experts and novices. In *Proceedings of the 15th international conference on Intelligent user interfaces* (pp. 329-332). ACM.

[14] Jansen, B. J., Spink, A., Blakely, C., and Koshman, S. 2007. Defining a session on web search engines. *Journal of The American Society for Information Science and Technology*, 58(6):862-871.

[15] White, R. W., and Morris, D. 2007. Investigating the querying and browsing behavior of advanced search engine users. In *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 255-262). ACM.

[16] 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* (pp. 21-30). ACM.

[17] Aula, A., Khan, R. M., and Guan, Z. 2010. How does search behavior change as search becomes more difficult?. In *Proceedings of the 28th international conference on Human factors in computing systems* (pp. 35-44). ACM.

[18] Bell, D. J., and Ruthven, I. 2004. Searcher's assessments of task complexity for web searching. In *Advances in Information Retrieval* (pp. 57-71). Springer Berlin Heidelberg.

[19] Broder, A. 2002. A taxonomy of web search. In *ACM Sigir forum* (Vol. 36, No. 2, pp. 3-10). ACM.

[20] Aula, A. 2003. Query Formulation in Web Information Search. In *ICWI* (pp. 403-410).