

# Towards Improving Web Search: A Large-Scale Exploratory Study of Selected Aspects of User Search Behavior\*

Hiroaki Ohshima, Adam Jatowt, Satoshi Oyama, Satoshi Nakamura,  
and Katsumi Tanaka

Graduate School of Informatics, Kyoto University  
Yoshida Honmachi, Sakyo, Kyoto 606-8501, Japan  
{ohshima, adam, oyama, nakamura, tanaka}@dl.kuis.kyoto-u.ac.jp

**Abstract.** Recently, the Web has made dramatic impact on our lives becoming for many people a main information source. We believe that the continuous study of user needs and their search behavior is a necessary key factor for a technology to be able to keep along with society changes. In this paper we report the results of a large scale online questionnaire conducted in order to investigate the ways in which users search the Web and the kinds of needs they have. We have analyzed the results based on the respondents' attributes such as age and gender. The findings should be considered as hypotheses for further systematic studies.

**Keywords:** Web search, search engines, user survey, user study.

## 1 Introduction

Although search engines are Web gateways, still, many times, it is difficult for users to satisfy their search needs. The reasons of lower effectiveness of Web search engines are multiple and may not always be immediately evident. We thus believe that the investigation of user search habits and the analysis of the needs they have is necessary and should be repeatedly conducted over time. This is especially important now in the view of the re-organization of the Web and the change in user focus and interest brought about by the Web 2.0 phenomenon and related technologies.

To this end we have decided to survey users for identifying the current usage patterns of search engines, related problems and necessary improvements that need to be done. The survey was administered on the group of 1000 online respondents in Japan in February 2008. We have divided users into equal or roughly equal groups considering their age and gender to analyze the results.

---

\* The complete report of this study is available at:  
[http://www.dl.kuis.kyoto-u.ac.jp/~ohshima/questionnaire2008\\_report.pdf](http://www.dl.kuis.kyoto-u.ac.jp/~ohshima/questionnaire2008_report.pdf)

The previous studies of the Web search (see [3] for survey) were often based on analyzing query logs. In [6] the authors manually classified a set of 500 queries from a query log into three basic classes: navigational, informational and resource, where each class could have some sub-classes. Since query logs contain large amounts of data, automatic approach was needed. Qiu and Cho [7] used click-through data and anchor-link distribution to correctly classify nearly 60% of queries into navigational and informational types. White and Morris [8] analyzed the searching and browsing behaviors of advanced searchers confirming that users who use advanced query syntax seem to be more efficient in search and spend more time online.

However when using anonymized query logs it is usually hard to determine the problems that average users have. This is because query and click-through data cannot always precisely answer questions on the degree to which the users had satisfied their search needs, the problems they encountered, neither their wishes or expectations they had. In addition, it is difficult to accurately interpret the data from the viewpoint of particular user attributes such as gender or age.

In this study we directly asked users about their activities related to the Web search and enhancements they would like. The objective was to make an overview of current user search activities on the Web and to identify potential problems and needs that users have when searching the Web. We have also focused on Web 2.0 sources.

Surveys made by Graphics, Visualization, & Usability Center (GVU) of Georgia Institute of Technology [4] were the first studies of the Web carried through online questionnaires. The objective of that project was to measure and describe the user population in the early years of the Web. More recently, Fogg et al. [1] reported results of an online questionnaire done in order to analyze user's perception of page credibility and to understand the key factors that influence it. The Pew Internet & American Life Project [5] is an initiative for continuous study of demographic aspects of Internet usage. Data collecting is usually done through telephone surveys. The topics of published surveys range from e-health, social communities to the impact of the Internet on politics. However, to the best of our knowledge, no similar questionnaire-based study that would directly concern the issues of Web search has been done so far. Lastly, with the advent of the so-called Web 2.0 it is becoming more appealing to analyze the activities of users in generating, utilizing and evaluating content on the Web.

## 2 Questionnaire Settings

The questionnaire was conducted online on a group of 1000 Web users in Japan between the 9th and 11th February 2008. The respondents volunteered for the study. For completing the survey the respondents received financial gratification. Subjects were grouped into equal size categories (250 respondents in each category) depending on their age: 20-29, 30-39, 40-49 and 50-59 years old. Also, in each category, half respondents were males and half females. Thus there was an even distribution of the four age groups for each gender. In addition, the subjects were selected so that the distribution of their living places roughly reflected the population distribution within Japan. In consequence, there were more users coming from densely populated areas (e.g., Tokyo and Osaka) than from the less populated ones (e.g., Ehime, Fukui).

We have presented the subjects with a multiple answers and top N answers' question types. The former gives the respondents the freedom to choose as many answers as they need among those provided, while the latter lets the subjects select N most significant answers. The questions and answers were written in Japanese. We show the translated results here.

We performed a chi-square significance test on the results to find significant differences between the answers of different user groups. For accuracy, we do not calculate it if answers to particular questions were chosen by less than the 5% of users.

### 3 Results

For each question, we first describe the aggregated results collected from all the participating users. Next, we will report statistically significant differences between particular user groups, provided there were any, in the form of an itemized list with the corresponding annotations: A (age) and G (gender). All the reported differences are with the significance strength  $p < 1\%$  unless stated otherwise. Due to space constraints we report here only some interesting findings.

#### Question on User search Activities

First, we asked a question about users' search activities on the Web via a multiple answers question (see Fig. 1). Not surprisingly, search by using standard search engines has been the most frequently chosen activity largely exceeding other choices. 93.7% users selected the Web search as one of their choices and 42.4% admitted issuing navigational search queries, that is, queries whose underlying intent is to return pointers to given pages or sites. Conventional search engines are also commonly used by users for news, image and video searches (33.3%, 22.9% and 14.8% of users, respectively). Only a relatively small number of users (about 5%) use advanced search operators such as "intitle:" or "-". This confirms the results from the previous study done by White and Morris [8]. We have also found out that many queries are issued from toolbars in Web browsers (37.5%). Also, relatively large number of subjects (14.9%) access links to copies of search results returned by search engines (e.g., when the original page cannot be accessed). This explains why major search engines provide access points to the cached copies of their search results despite risking violations of copyright laws.

- A The answer about the Web search using conventional search engines has been chosen at almost equal rates among different age groups. On the other hand, users in their 20s perform significantly more mobile search, video search, news search, image search, Q&A search and search in Wikipedia<sup>1</sup> than other age groups, and the decrease in the usage is inversely correlated to the age of users.
- G Searching for news and inside Wikipedia as well as searching using advanced operators are significantly more common among men than among women. On the other hand, women tend to more frequently search in Q&A sites ( $p < 5\%$ ).

---

<sup>1</sup> <http://www.wikipedia.org>

### Question on Favorite Content Types and User Activity on the Web

Next, we focused on the content types on pages that users view and create on the Web (Fig. 2). We asked users what content types they viewed in the last week and what content types they created or edited in the last month on the Web through a multiple

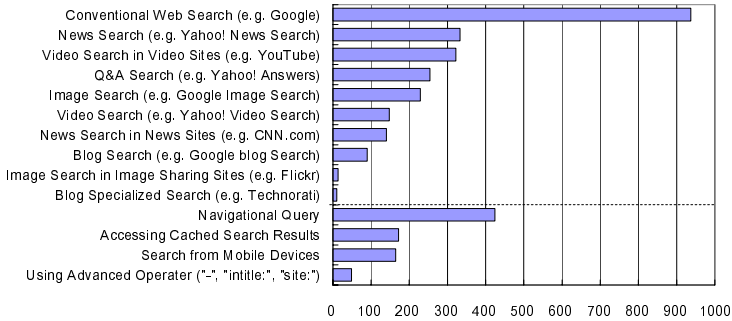


Fig. 1. Question on users’ searching activities on the Web (multiple answers)

answers’ question. News articles are the most frequently viewed content followed by blogs and videos from social sites like YouTube<sup>2</sup>. Wikipedia appears to be a highly popular destination (31.5%). Social bookmarking content in sites such as delicious<sup>3</sup> still remains in the interest of a rather small group of Web users. The low number of bookmarked and annotated pages appears to be one of the main problems preventing them to be used for search improvement on Web scale [2].

Considering user activity, it could be observed that majority of users do not perform any content generating/updating actions (49.5%). Actually, we have found out through additional question that the most popular activity is not social – “creating links in personal browsers” (30.6%). Users are usually less active in creating particular content type on the Web when compared to consuming it. We define an *activity indicator* of a given user group for a particular content type as the number of the users in the group that actively create the content divided by the number of the users in the group that only consume the content. If the activity indicator is 0 then the particular user group is just passively consuming the content, while the value 1 means all the users create, modify or distribute the content. Table 1 displays the activity indicators for different content types for all user groups. We can see that the activity indicator for Wikipedia is 0.08 (26/315) and 0.07 in case of video sharing sites for which only 29 users upload videos as opposed to 400 watching videos online. It is however much higher for blogs or public diaries, 0.47 (190/404). The activity indicator for blogs decreases along with the age of user groups.

A Users in older age groups view different content types significantly less frequently than the users in younger age groups (except for the questions about reading news in news sites or through conventional Web search engines). Users in their 20s are most active in almost each content generating/editing activities when compared to the other age groups. The exceptions here are updating own

<sup>2</sup> <http://www.youtube.com>

<sup>3</sup> <http://delicious.com/>

Web sites, writing private blogs, making bookmarks in browsers, editing Wikipedia content. Regarding the bookmarks in browsers, the young users probably just use search engines to re-find content or navigate to particular sites. They also use social bookmarking sites such as del.ici.ous more than older users. As for editing the Wikipedia content, the young respondents may lack knowledge and confidence to write or correct articles that are often of narrow focus and require certain knowledge.

- G Female users read statistically more blogs of their acquaintances, while male users read more news articles in news sites, read more Wikipedia content as well as watch more movies in video sharing sites. We also found out that men more frequently update content in Wikipedia than women do ( $p < 5\%$ ).

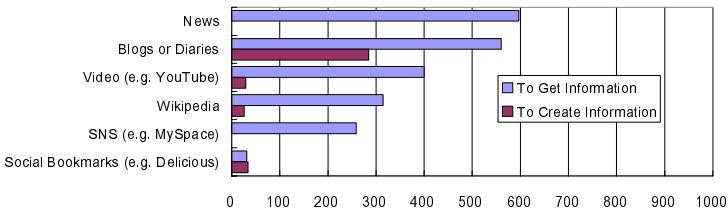


Fig. 2. Question about content types that users view and create/edit on the Web

Table 1. Activity indicators of different user groups for selected content types

Category	Sub-category	Videos	Wikipedia	Blogs
Age	20s	0.07	0.05	0.54
	30s	0.07	0.12	0.45
	40s	0.11	0.06	0.44
	50s	0.03	0.12	0.39
Gender	men	0.07	0.09	0.47
	women	0.08	0.07	0.47
<i>All</i>		<b>0.07</b>	<b>0.08</b>	<b>0.47</b>

### Question on Context of Search Results

We have also investigated what kinds of additional information users wish to have when receiving search results besides the usual data (titles, snippets, URLs) delivered by conventional Web search engines (Fig. 3). This was a top N question type in which the respondents had to select three results in the order of their importance. Interestingly, the most often selected choice is the one about displaying typical queries for which pages in search results are usually retrieved (55.7% of users have selected it as one of their top choices). Showing snapshots, images or graphs of returned pages is the second most popular answer (42.9%). No significant differences were found for different user groups.

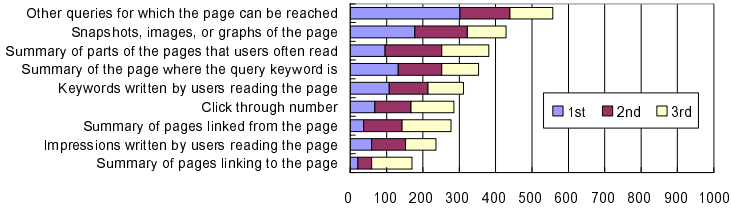


Fig. 3. Question on additional information users would like to have in Web search

**Questions on Types of Searched Objects and the Problems in Person Search**

We have also asked about what kinds of objects the users usually search for the information on the Web using multiple answers’ question (Fig. 4). We have found out that users search for any kinds of real-world objects. Especially, searches for the information about famous persons, places or products are the most common answers. Also, about 26.5% of users issue their own names into search engines in order to track the information about them on the Web. The Pew Internet & American Life Project reported the higher occurrence of this kind of search for the group of young Web users in USA (47% of US teens issue their own names as search queries [5]). In our study this number was 33% of the users in their 20s.

- A Young people submit their own names, names of other persons or their groups, whether famous or not, more frequently than the older users. On the other hand, users in their 30s submit more often names of places where they work(ed), live(ed) or learn(ed). Names of owned products are more often submitted as queries by users in their 30s and 40s than by the other age groups.
- G Women statistically more frequently than men submit names of places where they live, work or learn to search engines for receiving object-related information.

Through an additional question, we have also asked users about problems that they encounter when searching for the information on persons or small groups of persons on the Web. The trustworthiness of Web information appears to be the main problem for users. They often do not trust the information they encounter on the Web (51.9% of users). In addition, many subjects admitted that there is lots of untrue information on the Web (35.2%). Thus, there is an obvious need here for further improvements of search technologies related to personal information retrieval that would take into account the credibility of results apart from their relevance. In addition, the freshness of personal information is another aspect that the users cannot easily evaluate (46%).

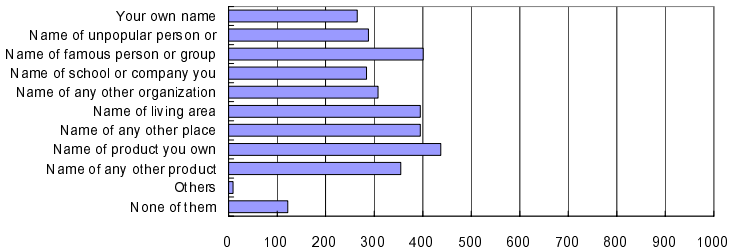


Fig. 4. Question on types of searched objects

## 4 Discussion

First, we briefly list some of the important results:

- Web search is very popular. When searching users often issue queries from browser toolbars, however, they rarely use advanced query operators. Cached copies of search results are relatively popular among searchers. Web search using search engines is done at the same rate independently of users' age, gender and locale. Men search and read news as well as Wikipedia content more frequently than women, while women read more blogs.
- News articles are highly popular and news reading or news searching are quite common activities. Wikipedia continues to be very popular, while social bookmarking sites still remain in the interest of few users. Younger Web users more frequently search and read various types of content on the Web when compared to older users. Women are more interested in blogs and diaries, while men prefer more Wikipedia and news content.
- Many users do not contribute to any content on the Web. The typical contribution is usually in the form of blog or diary comments. Consumers of blogs are thus more active than, for example, those of Wikipedia. The activity of users depends on age.
- Users would like to obtain information on the topics for which pages are popular on the Web. In general, there seems to be need for more additional data on search results to be provided for searchers.
- Person and product search are popular and users commonly search for information on current affiliations of persons or their reputation. Quite many users, especially the younger ones, check online information on themselves. Credibility of information and its freshness seem to be key problems in person search. Users often cannot obtain satisfactory information when searching information on persons. Person name disambiguation and credibility analysis of personal information should become important research areas.

Next we discuss some issues and implications:

- The above study has some weaknesses. First, it has been conducted within a single country making the results subject to certain cultural differences. Second, the selection of subjects was not purely random within the population, since the users volunteered for the study. Third, many problems could be only crudely identified since asking more detailed or complex questions would require presenting particular examples to users and providing them with detailed explanations.
- From the study, it appears that users require additional information on search results such as common queries issued by other users for which the results are returned. Although, the ranking algorithms applied by current search engines have been greatly improved, conventional interfaces for displaying search results seem to be still insufficient for users. Thus users have problems to find appropriate pages from results' lists using only the information on page titles and snippets. Therefore more additional information on returned pages should improve user Web search experience.
- The results of the group analysis can be used for improving personalized search and recommendation systems on the Web. For instance, female users may prefer to ob-

tain more search results from Q&A sites or blogs, while men from news sites and Wikipedia for their queries. Naturally, many users will prefer not to reveal their personal data and there are certain privacy issues here that should be considered. However, some user attributes can be actually automatically estimated to some extent. Even quite general user attributes such as age group, gender and location could already increase the effectiveness of Web search applications.

## 5 Conclusion

Concluding, Web search is the most common activity on the Web and it is important, although challenging, to continually measure users' satisfaction, recognize their behavioral patterns and problems they face. We believe that the effective studies should not only be done on numerical data, such as the one in query logs, but should also include direct questioning through which more complex issues can be highlighted and investigated (e.g., age, gender). To this end, we conducted an online survey on 1000 Web users focusing on general user activity on the Web.

## Acknowledgements

This research was supported in part by the National Institute of Information and Communications Technology, Japan, by Grants-in-Aid for Scientific Research (Nos. 18049041 and 18049073) from MEXT of Japan, and by the Kyoto University Global COE Program: Informatics Education and Research Center for Knowledge-Circulating Society.

## References

- [1] Fogg, B.J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M.: What makes Web sites credible?: a report on a large quantitative study. In: Proceedings of CHI 2001, pp. 61–68 (2001)
- [2] Heymann, P., Koutrika, G., Garcia-Molina, H.: Can social bookmarking improve web search? In: Proceedings of WSDM 2008, pp. 195–206 (2008)
- [3] Jansen, B.J., Pooch, U.: A review of Web searching studies and a framework for future research. *Journal of the American Society of Infor. Science and Techn.* 52(3), 235–246 (2001)
- [4] Kehoe, C.M., Pitkow, J.: Surveying the Territory: GVU's Five WWW User Surveys. *The World Wide Web Journal* 1(3), 77–84 (1996)
- [5] The Pew Internet & American Life Project, <http://www.pewinternet.org>
- [6] Rose, D.E., Levinson, D.: Understanding user goals in web search. In: Proceedings of the WWW 2004, pp. 13–19 (2004)
- [7] Qiu, F., Cho, J.: Automatic identification of user interest for personalized search. In: Proceedings of the WWW 2006, pp. 727–736 (2006)
- [8] White, R.W., Morris, D.: Investigating the querying and browsing behavior of advanced search engine users. In: Proceedings of SIGIR 2007, pp. 255–262 (2007)