The Electronic Library

A longitudinal study of information needs and search behaviors in science and technology: A query analysis

Minsoo Park Tae-Seok Lee

A longitudinal study of information needs and search behaviors in science and technology

A query analysis

Minsoo Park
Kangnam University, South Korea, and
Tae-Seok Lee
KISTI, South Korea

Received 2 April 2014
Revised 30 November 2014
Accepted 17 December 2014

Abstract

Purpose – This study aims at a longitudinal understanding of the user–system interactions from the context of science and technology at a query level.

Design/methodology/approach – The authors quantitatively analyzed log data sets culled from more than 24,820,416 queries submitted by users of a national scientific and technical information system, collected in 2008-2011.

Findings – In the fields of science and technology, the user search behaviors and patterns have remained stable. User queries are short and simple. In all, 80 per cent of the queries are made up of one-three terms. The length of query on a scholarly information system in the fields of science and technology is different from that of Web search. The former is longer than the latter. Search topics have shifted fast. "FUEL BATTERY", "NANO", "OLED", "CAR", "ROBOT" and "SMARTPHONE" were high-ranked queries from 2008 to 2011. It was found that the time to determine whether the users will stay on the site took about 10 seconds on average from the time of visit. If the users viewed the results of a list generated by the search query and took any action, such as detailed view, export or full-text download, most of them stayed more than 10 minutes on average.

Originality/value – Longitudinal user research using a query analysis helps to understand the information needs and behavioral patterns of users on information systems related to a specific field and those based on the Web. It also brings insights into the past, present and future events of a field. In other words, it plays a role as a mirror that reflects the flow of time. In the long run, it will be an historic asset. In the future, user studies using a query analysis need to be carried out from various (e.g. social, cultural or other academic disciplines) long-term perspectives on a continuous basis.

Keywords Data mining, Information needs, Query analysis, Science and technology, Search behaviors, User study

Paper type Research paper

Introduction

Since the advent of the Web, many studies have explored the interactions of users, information and technologies on the Web. Various tools that support Web search and access to many types and fields of information on the Web are available. Web search

The Electronic Library
Vol. 34 No. 1, 2016
pp. 83-98

0264-0473
DOI 10.1108/EL-04-2014-0058

engines, such as Google, Yahoo and Naver, are useful tools that people use to search for information. Researchers mainly use scholarly information systems, which offer information on the desired research area.

A contextual understanding of human information behavior can be obtained from a variety of measures, including query level, individual level, organizational level, social level and cultural level. User behavior research on Web search engines from the query aspect has been continuously conducted over the years. However, there is insufficient research on the behavior of specialized subject information systems. Therefore, this study aims at obtaining a longitudinal understanding of the user–system interactions from the context of science and technology at a query level.

The National Digital Science Library (NDSL) analyzed in this study is developed and run by Korea Institute of Science and Technology, which is a government research institute under the Ministry of Science, ICT and Future Planning. As a service platform for scientific and technical information, it provides high-quality information, such as that from journals, patents, industrial reports and trend analysis reports, and factual data to industrial and academic researchers.

The authors analyzed query data from the National Information System for Science and Technology. The data were collected for the period from 2008 to 2011. More than 24,820,416 queries and 3,165,712 sessions were analyzed quantitatively for this time period.

Literature review

The history of transaction log analysis goes back to the 1960s (Meister and Sullivan, 1967; Peters, 1993). Since then, log analysis has been a useful method to obtain a rich portrait of real behavior, to know the user and to improve the user experience. Log data are varied in terms of user characteristics, including searching strategy, language, knowledge level and tasks. Two major lines of research on log analysis include digital library systems and Web search engines.

Since the beginning of the 1980s, transaction log analysis has been widely used for understanding user behavior in online public access catalogues (OPACs) and digital library systems for statistical information of user–system interactions (Tolle, 1983; Jones *et al.*, 2000; Mahoui and Cunningham, 2000; Ke *et al.*, 2002). Jones *et al.* (2000) analyzed the log data from a digital library with technical reports from the field of computer science. In addition to an analysis of query terms, the use of search operators, searching errors and user demographics were discussed in their study. Ke *et al.* (2002) studied the general user behavior and usage of an e-journal system in the fields of science, technology and medicine by using a transaction log analysis.

Park *et al.* (2011) conducted a comprehensive study on information access and use in the fields of science and technology. Scientific and technical information were mainly accessed through work, Web sites and colleagues. Information needs were different in terms of the types of content. For example, users searching for research articles considered the availability of the original full-texts very important, whereas users mainly searching for patents and trend reports considered the availability of abstracts most important. Park and Lee (2013) studied the patterns and trends for the use of information systems in the fields of science and technology. More than 7,240,000 queries and 20,700,000 records for a full year were quantitatively analyzed. The study revealed

that the queries were short and simple (1.4 terms) and that user search sessions (8.2 queries) on the system were longer than on Web search engines.

Longitudinal Web studies from 1997 to 2001 showed that public Web searching is evolving. Search topics on the Web shifted from entertainment and sex to commerce and people, but that there has been little change in user query lengths (Spink *et al.*, 2002). Another longitudinal study by Piatetsky-Shapiro (2007) described the transition in data mining during the period from 1996 to 2005, showing the shift from university research to business analytics, becoming a part of the business applications.

A study by Spink and Jansen (2004) discussed Web search behaviors and trends at the human–computer interaction and query levels. According to this study, in general, Web queries are short and have a simple structure. People rarely used advanced search features. For a commercial Web search engine, a few terms were used with high frequency and user queries were diverse in terms of topic. A small number of users carried out longitudinal Web searching. Another longitudinal query study in the medical field reported that the information behaviors of medical professionals are more sophisticated than those of general search engine users, suggesting advanced query recommendations for assisting their information-seeking and retrieval process for health information (Yang *et al.*, 2011). Compared to those of Web search, users produced shorter queries, with general search terms, and used advanced search features more frequently while interacting with expert information systems (Fang *et al.*, 2011).

By tracing user activities recorded on logs, it is possible to grasp a detailed picture of the real-world behavior of users who use digital libraries and Web search to interact with systems and services during their information-seeking and retrieval process. A log analysis with a large sample size also allows opportunities for evaluation and improvement, as well as planning for new advanced services (Park, 2011; Park and Lee, 2013).

Data collection and analysis

As of September 2012, more than half-a-million (542,550) users registered with the NDSL and over three-and-a-half-million (3,550,527) full-texts were accessed. Most users came from research institutes:

- 39.9 per cent from enterprise research institutes;
- 35.5 per cent from institutions of higher education;
- 17.6 per cent from public research institutions; and
- 7 per cent from others.

Table I describes the current content (database) coverage in the fields of science, technology and medicine.

As part of a body of research studying the above-mentioned research problem, we analyzed data sets culled from more than 24,820,416 queries submitted by users collected between 2008 and 2011. The log data were analyzed at the levels of term, query and session (Park and Lee, 2013, p. 127):

> A *term* is any series of characters separated by a space; a *query* is the entire string of terms submitted by a searcher; and a *session* is the entire series of queries submitted by a user during one interaction with the information system with a time of not more than 30 minutes.

The search logs in this study were processed and analyzed, following the physical model of entity-relationship diagram (ERD) (Figure 1) (Park and Lee, 2013, pp. 127-128; Lee *et al.*, 2012): The criterion for the session analysis was a search log stream, where continuous mouse-clicks were performed within 30 minutes. The start and end times of the search session were saved in ss_searchsession, where the session analysis results were also saved. In addition, various events (number of search terms, number of searches, types of database accessed and more) occurring during one search session were also analyzed and recorded.

A Java program was used for analyzing the search log data sets. This longitudinal research is to understand user needs, search behaviors and topic trends in the area of science and technology from 2008 to 2011.

Results

Terms

Term frequency. Table II shows the top ten search terms, which are decided by term frequency in the log data set.

The most frequently entered terms were "LED", "SOLAR-BATTERY" and "CAR" from 2008 to 2010. These terms represent a strong user interest in the areas of Materials, Energy·Resources and Machinery. In 2011, "GRAPHENE" and "LED" were the top search terms, which were both categorized as New Materials. Most of the highly ranked terms are strongly related to the issues of Materials, Energy·Resources, Machinery and Life Science in 2008-2011.

Term co-occurrence. Table III represents the top ten co-occurrences for the NDSL data set during the period of 2008-2011. Term co-occurrence is useful, particularly to learn the specific usage of a term intended by a user outside the context of a specific query (Leydesdorff, 1989). Table III shows the top ten co-occurring term pairs for the information system.

In addition to "SOLAR CELL" maintaining the top spot for four years, the most frequently co-occurring term pairs in the queries are "STEM CELL" and "FUEL CELL" which are both associated with the issues of Life Science and Energy·Resources. The terms "SMARTPHONE" and "PHYSICAL THERAPY" are also highly ranked.

Queries

Queries per day and month. Figures 2 and 3 compare the patterns of queries per day and month from 2009 to 2011.

Table I.
NDSL contents (DB) (as of September 2012)

Type	Subtypes	Item no.
Journal articles	Domestic and international	54,274,813
Proceedings	Domestic and international	9,106,686
Theses	Domestic	1,338,799
Patents	Domestic, US, Japan, Europe, WIPO PCT	27,343,692
Reports	National R&D, industrial and technical analysis	177,756
Trends	Trend analysis reports	177,813
Standards	Korean Industrial Standards, ISO, IEC	62,555
Factual data	Chemistry, physics, human body, etc.	3,807,535
Total		96,289,649

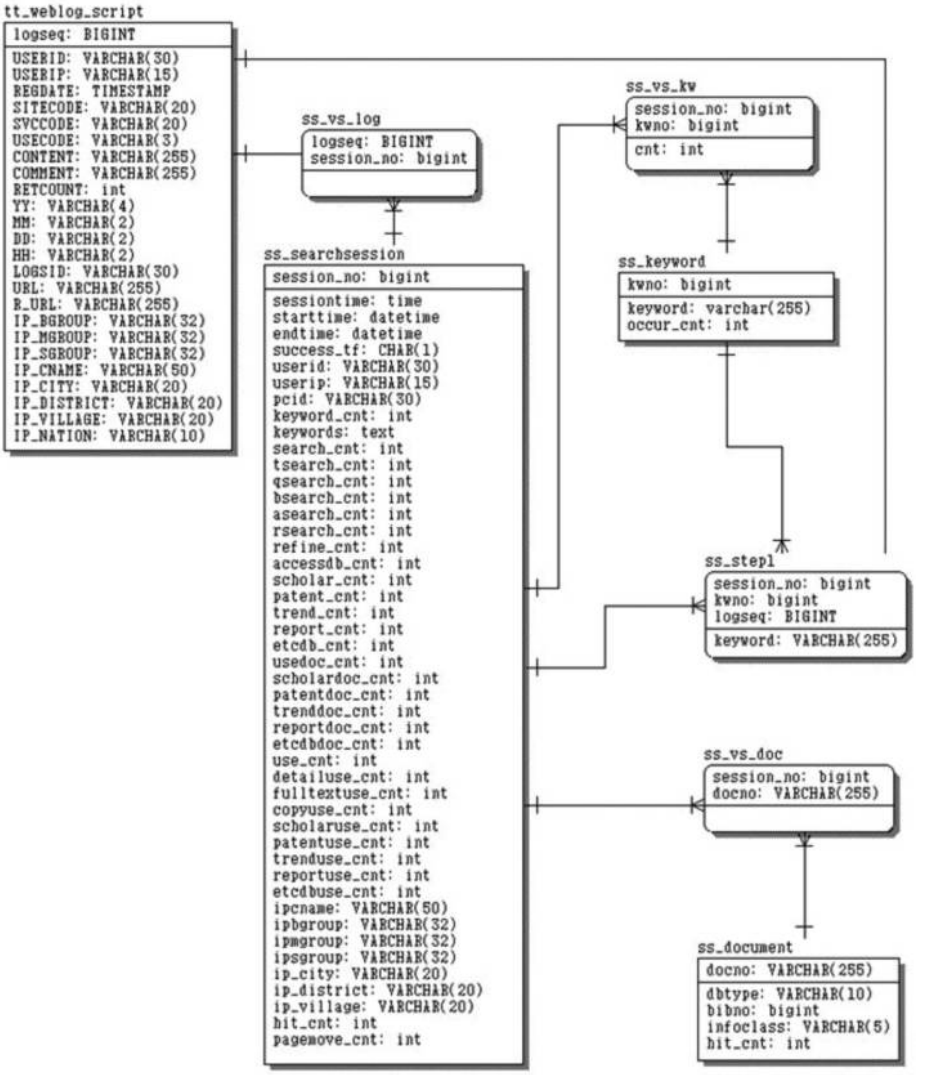

Figure 1.
Physical model of ERD

The mean queries per day were 1,112,403; 1,034,332; and 1,090,131, with an average number of 1,078,955 queries, for 2009, 2010 and 2011, respectively. On average, 88 per cent of the queries were entered during weekdays, with about 12 per cent queries submitted per day on weekends, showing a similar pattern of queries per day.

The mean queries per month were each 648,902; 603,360 and 635,910, with an average number of 629,390 queries, for 2009, 2010 and 2011, respectively. The highest queries were 855,813 in November 2009, 1,010,800 in November 2011 and 851,310 queries in March 2010, while the lowest ones were 289,914 in August 2009, 396,661 in December 2010 and 359,329 in February 2011. The early months in the year and the summer

Table II.
Top ten term frequency

	2008		2009		2010		2011	
Rank	Term	Frequency	Term	Frequency	Term	Frequency	Term	Frequency
1	LED	3,842	LED	19,961	LED	17,811	GRAPHENE	9,432
2	SOLAR-BATTERY	3,556	SOLAR-BATTERY	11,003	CAR	11,806	LED	9,168
3	ROBOT	3,487	FUEL-BATTERY	9,813	NANO	11,488	CAR	8,509
4	CAR	3,197	CAR	8,421	SOLAR-BATTERY	9,685	OLED	7,465
5	NANO	2,994	NANO	7,845	FUEL-BATTERY	8,309	ROBOT	7,273
6	FUEL-BATTERY	2,869	GINSENG	6,971	NANO-CELL	8,291	GINSENG	7,152
7	RFID	2,075	WIND-POWER	5,712	ROBOT	6,201	KIMCHI	6,758
8	OLED	1,965	RFID	5,698	WIND-POWER	6,110	SMARTPHONE	5,769
9	CHILD DEVELOPMENT	1,849	SOLAR CELL	5,488	MICROBE	5,731	SOLAR-BATTERY	5,073
10	SENSOR	1,693	WATER	5,417	CARBON-NANOTUBES	5,671	MICROBE	5,059

Table III.
Top ten term co-occurrences

Rank	2008	2009	2010	2011
1	SOLAR CELL	SOLAR CELL	SOLAR CELL	SOLAR CELL
2	PHYSICAL THERAPY	FUEL CELL	RISK ASSESSMENT	SMART PHONE
3	AEROBIC EXERCISE	STEM CELL	STEM CELL	CONCURRENT ENGINEERING
4	X-RAY DIFFRACTION	CARBON NANOTUBE	FUEL CELL	REGULATORY FOCUS
5	PLATE HEAT EXCHANGER	OCCUPATIONAL THERAPY	CARBON NANOTUBE	SENSORY INTEGRATION
6	FUEL CELL	CEREBRAL PALSY	ALUMINUM ALLOY WELDING	PULMONARY FUNCTION
7	LITHIUM ION BATTERY	MICROBIAL FUEL CELL	SMART GRID	HEPATIC FUNCTION
8	CARBON DIOXIDE	SMART GRID	CARBON NANOTUBES	RENAL FUNCTION
9	ASSET ALLOCATION	LIQUID CRYSTAL	OCCUPATIONAL THERAPY	SPLEEN FUNCTION
10	LACTIC ACID	HEAT TRANSFER	FLAME RETARDANT	STOMACH FUNCTION

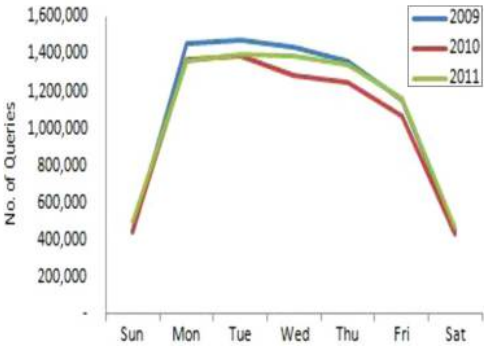

Figure 2.
Queries per day

Figure 3.
Queries per month

months (particularly August) showed a low tendency for use, while March, April and November had a relatively high pattern of use.

Query length. Figure 4 shows the result of comparison and analysis of query length by year. The term "query length" refers to the number of terms per query. As you can see from the figure, queries are generally short and contain few terms from 2008 to 2011. Of the total queries, 70-80 per cent of queries are made up of one to three terms.

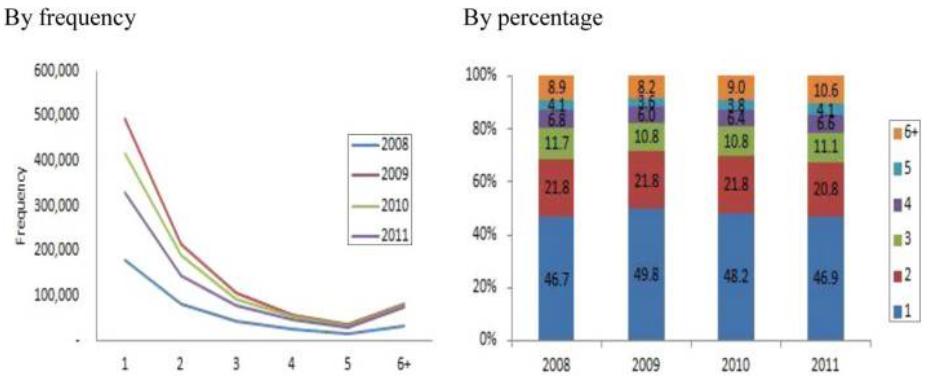

Figure 4.
Query length

Most frequent queries. Table IV represents the most frequent queries in NDSL from 2008 to 2011.

The most frequently used queries were "SOLAR BATTERY" in 2008, "LED" in 2009 and 2010 and "GINSENG" in 2011. "FUEL BATTERY", "NANO", "OLED" (which refers to organic light-emitting diode), "CAR", "ROBOT" and "SMARTPHONE" were also high-ranked queries from 2008 to 2011. These queries represent the users' high information needs on the issues of Energy Resources, Materials, Health Medicine, Life Science and Machinery.

Sessions

Sessions per day and month. Figures 5 and 6 show the number of sessions per day and month with a comparative analysis for the period of 2009 to 2011.

The mean number of sessions per day was 307,071 with weekdays showing a higher number of sessions in 2009-2011. The average number of sessions per month in 2009-2011 was 182,384. When the pattern of the monthly average number of sessions was examined, March ranked highest with 231,293 sessions and February and December ranked lowest, each with 124,159 sessions and 169,667 sessions, respectively.

Session length. Figure 7 shows the range of the session length. The session length was obtained by calculating the number of queries per session.

Of the total NDSL users, 36-47 per cent of them entered one or two queries during their session. In 2008, almost one in two NDSL users entered one or two queries. In 2008-2010, one out of ten users entered three queries during a search session. Only 5 per cent of the users in 2011 entered three queries. In general, one or two queries were used to formulate their information needs.

Session duration. In this study, a search session is the same as search visits, and generated a session based on a 30-minute session timeout. It is counted as one session if it is entered after initial access and has less than 30 minutes of click stream. Table V shows the aggregate statistics for session durations.

On this information system, most of the search sessions were carried out within 30 minutes. For the period 2008-2011, the average statistics of a search session by users were as follows: a mean of 24 minutes, a minimum of one second, a maximum of 14 hours and a mode of 11 seconds. Most of the users made a quick decision regarding whether to stay (use the information system) or not, taking about 9-13 seconds from the moment the user entered the site.

Figure 8 shows a comparison of the analyzed data by year for the period 2008-2011 on the session duration of the search success/failure rate. *Search success* means any action taken, such as detailed view, exporting or downloading full-texts, after the user viewed the search result list generated by the entered query. On the other hand, a *failed search* is a case where no additional action is taken after the search result list is displayed. In this case, in general, the query is reformulated or the user leaves the site.

About 30 per cent of the sessions were less than one minute in duration with a low search success rate and a high search failure. Approximately 25 per cent of the sessions were between one and five min, having a transition point in terms of search success and failure (high search success rate was between 50.5 and 56.8 per cent, and low failure rate between 43.1 and 49.5 per cent). The sessions of one minute or less had a high failure search rate, while that of 10 minutes or more showed a high search success rate.

Table IV.
Most frequent queries

	2008		2009		2010		2011	
Rank	Queries	Frequency	Queries	Frequency	Queries	Frequency	Queries	Frequency
1	SOLAR BATTERY	2,614	LED	8,121	LED	5,900	GINSENG	4,460
2	FUEL BATTERY	2,039	SOLAR BATTERY	7,017	SOLAR BATTERY	4,030	ROBOT	4,357
3	LED	2,015	FUEL BATTERY	6,617	FUEL BATTERY	3,639	SMARTPHONE	4,262
4	CAR	1,727	WIND POWER	4,014	NANO	3,553	OLED	4,118
5	NANO	1,597	GINSENG	3,937	ELECTRIC VEHICLE	2,747	CAR	3,929
6	RFID	1,059	SOLAR ENERGY	3,470	GRAPHENE	2,686	KIMCH	3,893
7	OLED	959	NANO	3,439	WIND POWER	2,663	GRAPHENE	3,882
8	SENSOR	783	ATOPY	3,324	ROBOT	2,583	SCHIZANDRA	3,509
9	ROBOT	760	OBESITY	3,108	CAR	2,523	LED	3,157
10	CARBON NANO TUBES	688	CARBON NANO TUBES	2,983	SOLAR ENERGY	2,437	SOLAR BATTERY	2,969

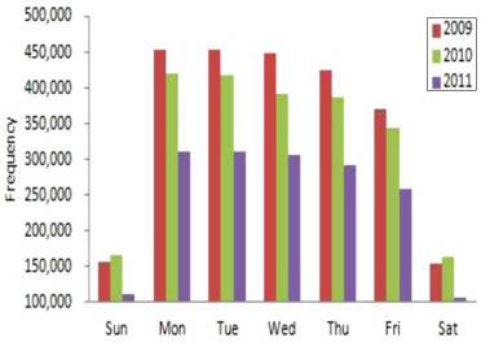

Figure 5.
Sessions per day

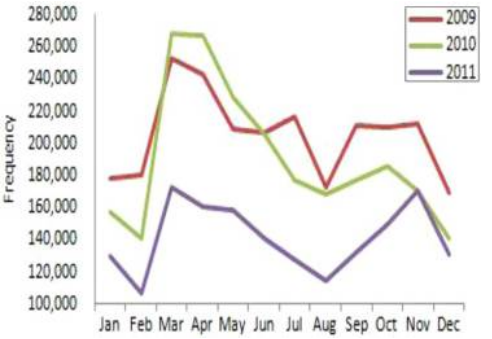

Figure 6.
Sessions per month

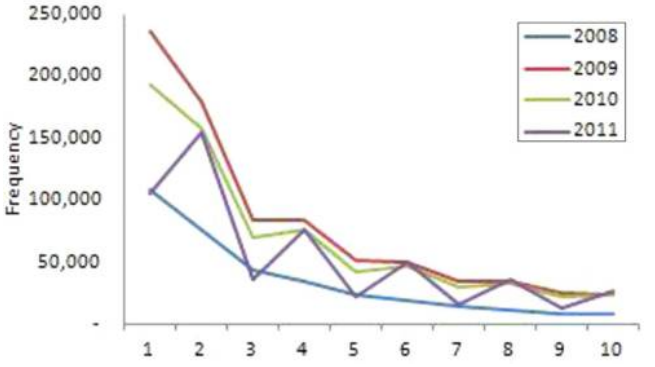

Figure 7.
Session length

Statistics	2008	2009	2010	2011
Mean	0:24:19	0:26:07	0:24:15	0:23:26
Minimum	0:00:01	0:00:01	0:00:01	0:00:01
Maximum	9:14:40	11:15:10	10:10:07	14:28:12
Mode	0:00:13	0:00:11	0:00:10	0:00:09

Table V.
Statistics for NDSL sessions

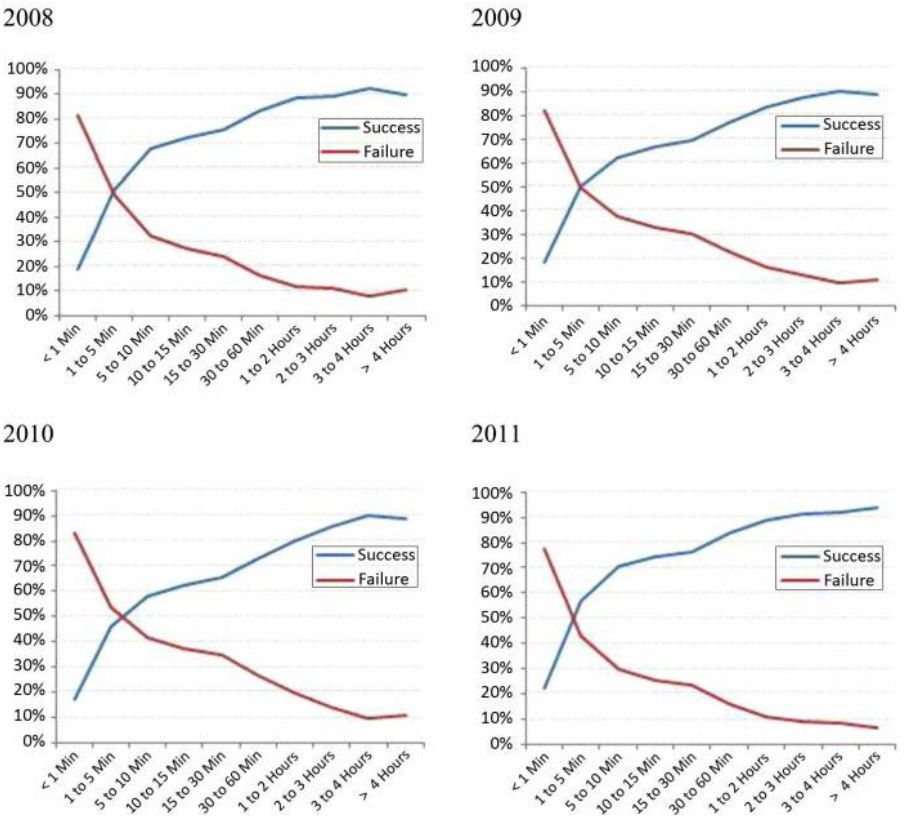

Figure 8.
Session duration and search success/ failure by year

Session topic classification

Figure 9 shows the results of the topic classification of the sessions. The classification consists of a total of 18 science and technology topic categories developed by the Ministry of Science, ICT, and Future Planning and the Korea Institute of S&T Evaluation and Planning in 2005.

Aerospace·Astronomy·Marine (2008), Environment (2009) and Life Science (2010) ranked first, each with 17.4, 20.8 and 16.1 per cent, respectively. In 2008, one in three queries submitted to NDSL were related to Aerospace·Astronomy·Marine (17.4 per cent) and Energy·Resources (14.9 per cent). In 2009, Environment (20.8 per cent) and Information (17.8 per cent) were popular topics from the fields of science and technology. In 2010, one in three queries submitted to NDSL were related to Life Science (16.1 per cent) and Information (14.9 per cent).

Characteristics of user search. Figure 10 shows the results of a comparative analysis of the search behavior pattern by search type. *Basic search* is a field-specific, itemized search (such as title and abstract). *Integrated search* is a simple search of all NDSL database keywords. *Quick search* is a simple keyword search by content. *Refined search* refers to a limited search by title or scope of year from the search results.

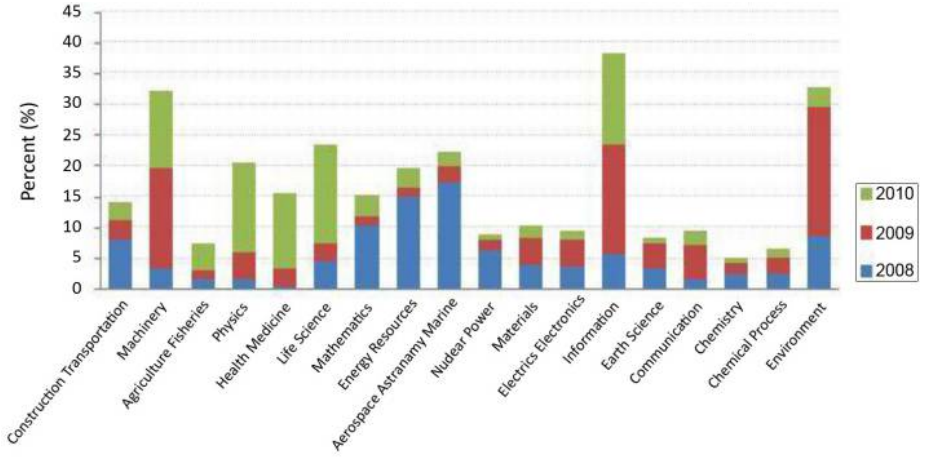

Figure 9.
Topic trends by year

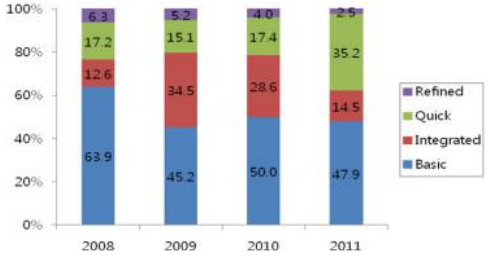

Figure 10.
Search type by year

From 2008 to 2011, users searched NDSL through the feature of basic search most, followed by integrated search and quick search. Refined search was used rarely throughout the years (Table VI).

As the number of keywords per query is 1.3 on average, there has been no significant change in four years. As the number of queries per session is 8.0 on average, it shows a consistent growth pattern from 2008 to 2011: from 5.5 in 2008 to 10.5 in 2011.

Discussion and conclusion

The major findings of this study are as follows.

10 seconds or 10 minutes: users are impatient

It is a well-known fact that Web users lack patience. The studies on behaviors of Web users from Nielsen (1993, 1999) show the attributes of information users on the Web. According to

Statistics	2008	2009	2010	2011
Average number of terms in a query	1.2	1.3	1.4	1.2
Average number of queries in a session	5.5	7.7	8.2	10.5

Table VI.
Average number of terms in a query and queries in a session

Nielsen, in particular, Web users lack patience on a site whose design is difficult to understand or on one that is slow. Users are reluctant to wait or learn how to use the site. After visiting a Web site, they want to use all the functions within a few seconds. Otherwise, they will simply leave the site and visit other sites to meet their requirements.

Our research supports these results. It was found that the time to determine whether an NDSL user will stay on the site took about 10 seconds, on average, from the time of visit. It is possible that the user could not find the desired information, or he/she was not really interested in the displayed information (it was not what he/she was really looking for), or it was a wrong (accidental) visit. On the other hand, if the users viewed the results of a list generated by the search query and took any action, such as detailed view, export or full-text download, most of them stayed more than 10 minutes, on average.

If the above is true, what can a system designer do at the crossroads between 10 seconds and 10 minutes? The target user group of the NDSL is clear. It consists of research and development researchers from the fields of science, technology and medicine. The authors suggest that an advanced information service is needed to satisfy user information needs, such as a personalized service or a recommendation service using semantic technology. Further, for analyzing the pattern and behavior of a user's information search, it is imperative to make the most of user experience by supplying a user-friendly interface with intuitive aesthetics and convenient tools.

User queries are short and simple, but search sessions are longer

In the fields of science, technology and medicine, the user search behaviors and patterns have remained stable. For NDSL, user queries are short and simple. Eighty per cent of the queries are made of one to three terms. This is similar to the results of Web search research conducted using a log analysis (Montgomery and Faloutsos, 2001; Piatetsky-Shapiro, 2007; Spink *et al.*, 2002; Spink and Jansen, 2004). Most users frequently conduct a Web search using queries in the months of November and March and on weekdays as compared to weekends. The occurrence of sessions also shows the same pattern. With respect to the search function, a simple or integrated search is preferred to an advanced search.

It has been found that the length of query on a scholarly information system in the fields of science and technology is different from that of Web searching. The former is longer than the latter. Another noticeable point from the results of the query length analysis is that the query length increased steadily from 2008 to 2011 (from 5.5 to 10.5 minutes).

The NDSL's user group is primarily made up of researchers from research institutes, professors and graduate or doctoral students at universities from the fields of science, technology and medicine. In other words, it can be interpreted that this user group has more detailed and professional information needs than general users on the Web (Park *et al.*, 2011; Yang *et al.*, 2011). It is also shown that a frequent occurrence of query reformulation is one of the reasons that the query length is so long. For them, there is a need to provide a query recommendation function and the query history features. Further studies of query reformulation are required.

Search behaviors unchanged, search topics shifted fast

The most frequently entered terms were "LED", "SOLAR-BATTERY" and "CAR" from 2008 to 2010. These terms represent a strong user interest in the areas of Materials, Energy·Resources and Machinery. In 2011, "GRAPHENE" and "LED" were the top

search terms, which were both categorized as New Materials. Most of the highly ranked terms were strongly related to the issues of Materials, Energy·Resources, Machinery and Life Science in 2008-2011.

The most frequently used queries were "SOLAR BATTERY" in 2008, "LED" in 2009 and 2010 and "GINSENG" in 2011. "FUEL BATTERY" "NANO", "OLED", "CAR", "ROBOT" and "SMARTPHONE" were also high ranked queries from 2008 to 2011. These queries represent the users' high information needs on the issues of Energy·Resources, Materials, Health·Medicine, Life Science and Machinery.

Aerospace·Astronomy·Marine (2008), Environment (2009) and Life Science (2010) ranked first, each with 17.4, 20.8 and 16.1 per cent, respectively. In 2008, one in three queries submitted to NDSL were related to Aerospace·Astronomy·Marine (17.4 per cent) and Energy·Resources (14.9 per cent). In 2009, Environment (20.8 per cent) and Information (17.8 per cent) were popular topics in the fields of science and technology. In 2010, one in three queries submitted to NDSL were related to Life Science (16.1 per cent).

In Korea, a recent trend in the fields of science, technology and medicine is that the research and development in these fields have expanded to promote interdisciplinary research in the fields of biotechnology, nanotechnology and robotics. For example, a major trend of topics related to nanomaterials, such as nanotechnology and nanoscience and technology, include the following:

- applications of griffin and energy conversion of biomaterials applications of green energy and solar cells; and
- emergence and clinical applications of nanoparticles to the diagnosis and treatment of cancer.

With this contextual understanding, the results of this study clearly show the current hot trends are in the fields of science, technology and medicine in Korea.

For the users of scientific and technical information, the search behaviors remained unchanged and the information needs/topics shifted. The longitudinal query analyzes and research on the Web and in the specific area of information systems show that there is a tendency that human information behavior does not change easily. On the other hand, the users' information needs in the fast-developing fields of science and technology change dynamically. A user-oriented information service/system has to be designed along with sustainable content and service development which satisfies the current and potential users' needs and behaviors.

As shown by the results of this study, longitudinal user research using a query analysis helps to understand the information needs and behavioral patterns of users on information systems related to a specific field and those based on the Web. It also brings insights into the past, present and future events of a field. In other words, it plays a role as a mirror that reflects the flow of time. In the long run, it will be an historic asset. In the future, user studies using a query analysis need to be carried out from various long-term perspectives, such as academic or social, on a continuous basis.

References

Fang, Y., Somasundaram, N., Si, L., Ko, J. and Mathur, A.P. (2011), "Analysis of an expert search query log", *SIGIR 2011*, Beijing, 24-28 July.

Jones, S., Cunningham, S.J., McNab, R. and Boddie, S. (2000), "A transaction log analysis of a digital library", *International Journal on Digital Libraries*, Vol. 3 No. 2, pp. 152-169.

Ke, H.-R., Kwakkelarr, R., Tai, Y.-M. and Chen, L.-C. (2002), "Exploring behavior of e-journal users in science and technology: transaction log analysis of Elsevier's ScienceDirect Onsite in Taiwan", *Library & Information Science Research*, Vol. 24 No. 3, pp. 265-291.

Lee, T., Jeong, D., Moon, Y., Park, M. and Hyun, M. (2012), "An analytic study on the categorization of query through automatic term classification", *Journal of Korea Information Processing Society*, Vol. 19 No. 2, pp. 133-138.

Leydesdorff, L. (1989), "Words and co-words as indicators of intellectual organization", *Research Policy*, Vol. 18 No. 4, pp. 209-223.

Mahoui, M. and Cunningham, S.J. (2000), "A comparative transaction log analysis of two computing collections", *Research and Advanced Technology for Digital Libraries, Lecture Notes in Computer Science*, Vol. 1923, pp. 418-423.

Meister, D. and Sullivan, D.J. (1967), "Evaluation of user reactions to a prototype on-line information retrieval system", Report to NASA by the Bunker-Ramo Corporation under Contract No. NASA-1369, Report No. NASA CR-918, ERIC ED 019 094, Canoga Park, CA.

Montgomery, A. and Faloutsos, C. (2001), "Identifying web browsing trends and patterns", *IEEE Computer*, Vol. 34 No. 7, pp. 94-95.

Nielsen, J. (1993), *Usability Engineering*, Academic Press, Boston, MA.

Nielsen, J. (1999), *Designing Web Usability: The Practice of Simplicity*, New Riders, Indianapolis, IN.

Park, M. (2011), "User participation evaluation of a scholarly information site", *Journal of the Korean Society for Information Management*, Vol. 28 No. 4, pp. 85-97.

Park, M., Lee, S., Choi, H. and Jung, J. (2011), *A Study on Users' Needs and Behaviors of Science and Technology Information in Korea*, KISTI, Seoul.

Park, M. and Lee, T. (2013), "Understanding science and technology information users through transaction log analysis", *Library Hi Tech*, Vol. 31 No. 1, pp. 123-140.

Peters, T. (1993), "The history and development of transaction log analysis", *Library Hi Tech*, Vol. 42 No. 11, pp. 41-66.

Piatetsky-Shapiro, G. (2007), "Data mining and knowledge discovery 1996 to 2005: overcoming the hype and moving from university to business and analytics", *Data Mining and Knowledge Discovery*, Vol. 15 No. 1, pp. 99-105.

Spink, A. and Jansen, B.J. (2004), *Web Search: Public Searching of the Web*, Kluwer, New York, NY.

Spink, A., Jansen, B.J., Wolfram, D. and Saracevic, T. (2002), "From e-sex to e-commerce: web search changes", *IEEE Computer*, Vol. 35 No. 3, pp. 107-111.

Tolle, J.E. (1983), *Current Utilization of Online Catalogs: Transaction Log Analysis, Final Report to the Council on Library Resources (Volume 1)*, Research Report No. OCLC/OPR/RR-83/2, OCLC, Dublin, OH.

Yang, L., Mei, Q., Zheng, K. and Hanauer, D.A. (2011), "Query log analysis of an electronic health record search engine", *AMIA Annual Symposium Proceedings*, Vol. 2011, pp. 915-924.

Corresponding author
Minsoo Park can be contacted at: mspark7@gmail.com