

U.S. VERSUS EUROPEAN WEB SEARCHING TRENDS

Amanda Spink*

School of Information Sciences and Technology
The Pennsylvania State University
004C Thomas Building, University Park PA 16802
Tel: (814) 865-4454 Fax: (814) 865-5604
E-mail: spink@ist.psu.edu

Seda Ozmutlu & Huseyin C. Ozmutlu
Department of Industrial Engineering
Uludag University
Gorukle Kampusu, Bursa, 16059, Turkey
Tel: (90-224) 442-8176 Fax: (90-224) 442-8021
E-mail: seda@uludag.edu.tr

Bernard J Jansen
School of Information Sciences and Technology
The Pennsylvania State University
E-mail: jjansen@ist.psu.edu

* To whom all correspondence should be directed.

INTRODUCTION

As the Web is becoming a worldwide phenomenon we need to understand what searching trends are emerging across different global regions. Are there regional differences in Web searching? What are the differences between searching by the United States population compared to Europeans? As part of a body of research studying these questions, we have analyzed two data sets culled from more than one million queries submitted by more than 200,000 users of the Excite Web search engine collected in May 2001 and the FAST Web search engine (All theWeb.com), collected in February 2001.

We compare the searching behavior of largely European FAST Web search engine users (mostly German) with Excite Web search engine users who are largely U.S. This comparative study shows differences in Web searching by U.S. and European users. Specifically, the results suggest some differences in the topics searched and searching behaviors.

SCOPE OF STUDY

Excite (<http://www.excite.com>) is a major Internet media company offering Web searching and a personalization portal. FAST (<http://www.fast.no>) and (<http://www.alltheweb.com>) Web search engines. We analyzed a data set from the FAST search engine Web queries submitted on February 6, 2001. The entries are given in the

order they arrive. New sessions/users are identified through a user ID and each query is given time stamps in hours, minutes and seconds. Previous studies show that approximately 84% of Excite users were located in the United States (Spink, Bateman & Jansen, 1998). The majority of FAST users are believed by FAST to be from Europe, mostly from Germany. The specifications of the Excite and FAST query dataset are listed in Table 1

Our study is limited to the analysis of user queries, as we had no access to data on the Web sites users' accessed. While only data from Excite and FAST was examined, we provide a baseline for comparing Web searching trends in the United States and Europe. The data analysis in this study has the following parts: (1) session and query length and structure, (2) mean query and session durations, (3) search terms per query, and (4) terms in queries. The multimedia queries were also sifted from the entire query logs.

Each query log record contained three fields: *Identification*: anonymous code assigned by Web company server to a user machine, *Time of Day*: in hours, minutes, and seconds, *Query*: user terms as entered. We analyzed user: *Sessions* - entire query sequence by a user; *Queries* - one or more entered terms; *Terms* - any string of characters bounded by white space.

Table 1. Excite and FAST 2001 data sets.

2001 Data Set	Sessions	Queries	Terms
Excite	262,025	1,025,910	1,538,120
FAST	153,297	451,551	1,350,619

RESULTS

Are there global regional differences in Web searching? Table 2 compares results from the FAST and Excite query analysis.

Table 2. Comparative statistics for FAST and Excite 2001 Web query data sets – one million queries per study.

Variables	2001 Excite Study (1.2M Queries)	2001 FAST Study (1.2M Queries)
Mean Terms Per Query	2.6	2.3
Terms Per Query		
1 Term	26.9%	25%
2 Terms	30.5%	36%
3+ Terms	42.6%	39%
Mean Queries Per Session	2.3	2.9
Session Size		
1 Query	55.4%	53%
2 Queries	19.3%	18.9%
3+ Queries	25.3%	29%

Variables	2001 Excite Study (1.2M Queries)	2001 FAST Study (1.2M Queries)
Mean Pages Viewed Per Query	1.7	2.2
% of Use of 100 Most Frequently Occurring Query Terms	22%	14%

There is little difference in the mean terms per query for FAST and Excite users. The previous Excite studies also showed little change in Web query length from 1997 to 2001 (Spink, Jansen, Wolfram & Saracevic, 2002), and this trend continues with FAST users. Web queries by both sets of users are still relatively short and limited to few words. The continuing use of one short simple query by users suggests that targeting toward specific, high frequency words (such as free, sex, games, weather, maps, etc.) by Web information content providers is critical in order to reach Web users.

Query Sessions

FAST users generated slightly more queries per session than Excite user sessions. The mean length of Excite user sessions has not changed significantly from 1997 to 2001 (Spink, Jansen, Wolfram & Saracevic, 2002). The *Alta Vista study* reports also on queries per session - 77.6% of sessions had one query, 13.5% had two, and 4.4% had three (Silverstein, Henzinger, Marais, & Moricz, 1999). Most users in the Excite and Alta Vista studies entered only one query. The mean queries per FAST user sessions were slightly longer than of Excite users. In addition, a slightly higher proportion of FAST users submitted three or more unique queries. However, users typically do not add or delete many terms in their subsequent queries. Comparing data from Excite in 1999 and FAST, we see that some FAST users generated longer queries and sessions by duration (in seconds). Many FAST user sessions also involved queries related to multiple topics during the same search or multitasking. Multitasking Excite sessions often include between 2 to more than 10 topics with a mean of 2 topic changes per session (Spink, Ozmutlu & Ozmutlu, 2002).

FAST users viewed more retrieved pages per query than Excite users. Since 1997, Excite users were viewing fewer pages of results per query. Were FAST users more willing to view more pages of results? Were Excite users more satisfied with the results and did not need to view more pages? The trend toward viewing an increasing number of pages, combined with the larger queries per session, suggests that FAST users may be more interactive during their Web searching sessions and more persistent in their Web searching sessions.

Previous studies show that Excite users have a low tolerance for wading through large retrievals (Spink, Jansen, Wolfram & Saracevic, 2002). Another recent longitudinal study of 20,000 Internet users also concluded that Web searching is changing little in session times (Montgomery & Faloutsos, 2001). FAST users appear to be countering the trend toward greater simplicity in Web searching. We see higher levels of interactivity by FAST users, as they create more queries per session and more pages viewed per query. The variation in level of user interactivity is an area for further research.

Web Query Topics

Are there differences in topics searched by U.S. and European users? We classified a random sample of approximately 2,500 Excite and FAST (English language) queries, into 11 non-mutually exclusive, general topic categories.

Table 3. Comparison of general topic categories for Excite and FAST.

Rank	2001 Excite Data Set (2453 queries)	2001 FAST Data Set (2503 English Queries)
1	24.7% Commerce, travel, employment or economy	22.5% People, places or things
2	19.7% People, places or things	21.8% Computers or Internet
3	11.3% Unknown	12.3% Commerce, travel, employment, or economy
4	9.6% Computers or Internet	10.8% Sex or pornography
5	8.5% Sex or pornography	9.1% Entertainment or recreation
6	7.5% Health or sciences	7.8% Health or sciences
7	6.6% Entertainment or recreation	4.8% Society, culture, ethnicity or religion
8	4.5% Education or humanities	4.7% Performing or fine arts
9	3.9% Society, culture, ethnicity or religion	2.9% Education or humanities
10	2% Government	2.7% Government
11	1.1% Performing or fine arts	0.6% Unknown

The major categories of FAST search topics were *People, places and things*, and *Computers or Internet*. FAST users were searching less on e-commerce related issues and more on people and computer related issues. The major Excite categories such *Commerce, travel, employment or economy* and *People, places or things* were more highly ranked than *Sex, pornography or preferences* and *Entertainment or recreation*. The pre-dominance of e-commerce in Excite queries coincided with changes in information distribution on the publicly indexed Web, greater use of e-commerce by U.S.-based companies (Spink, Jansen, Wolfram & Saracevic, 2002). By 2001, Excite Web searching continues to move from an *entertainment to a business medium*.

Table 4 shows the top 20 terms by frequency for FAST and Excite 2001 users.

Table 4. Top 25 terms for Excite and FAST 2001.

Top 20 Excite 2001 Terms	Top 20 FAST 2001 Terms
and	and
of	free
free	download
the	sex
sex	com
in	pictures
christmas	uk
nude	nude
for	mp3
pictures	hotel
new	windows
pics	pics
to	www
a	jpg
music	crack
games	software
stories	index
or	education
woman	new
university	resume

FAST users entered a lower number of high frequency terms than Excite users. The top 20 FAST query terms by frequency formed a smaller proportion of all terms than we see with Excite users. The vocabulary of FAST users seems to be broader than previously identified for Web users (Spink, Jansen, Wolfram & Saracevic, 2002). For both FAST and Excite users, despite some high frequency terms (e.g., sex, and, etc), a large number of terms are not repeated or used with low frequency, including personal names, number of spelling errors, non English terms, Web specific terms (e.g., URLs).

The findings suggest differences in search topics for FAST and Excite users. For Excite users, the previous studies (Spink, Jansen, Wolfram & Saracevic, 2002) showed an on-going shift in Web users' search topics. This finding suggests some differences in Web searching between European and U.S. based users in search length and topics.

DISCUSSION

Our research confirms previous studies that suggested little movement towards longer or more complex user sessions (Spink, Jansen, Wolfram & Saracevic, 2002). Our comparative study of Web query data from two major commercial search engines provides large-scale findings related to trends and differences in Web searching on a

global level. The paper also suggests interesting differences in search behavior and topics searched by U.S and European users.

The study suggests some potential regional differences in public Web. Even though the size of the queries tends to stay the same, there is some movement towards more frequent queries with FAST users. Search topics continue to be a key vehicle of change. Overall, we see some differences and similarities in FAST and Excite users related to queries, sessions and terms. Why are FAST users slightly more interactive? We can speculate that greater interactivity may be linked to possible cultural differences in information behaviors and the structure of the Web interface offered by each search engine. The relationship between interactivity and search effectiveness is also not possible to assess using the query data we examined and compared. It is also not unusual to see Web search query sessions that consist of multiple topics or multitasking.

The nature of searching is changing and the search topics are fluid. Why are U.S users more focused on e-commerce search topics? Again, we can speculate that the differences in search topics identified in this comparative study are related to the greater impact of e-commerce on the U.S economy. Since 1997, the impact of a shift to e-commerce has been a major economic and social shift in the U.S. It seems that some European users may be more willing to approach Web searching as a more interactive and longitudinal process. Alternatively, Europeans are not using the Web as much for business information, but more for socially related information.

These possible cultural and social differences represent a major challenge for both Web search engines and the people who use them. We may see a greater shift to e-commerce search topics by European users in the future as more European businesses move towards the Web. People are spending more and more time creating, seeking and using electronic information. However, their interactions with Web search engines are still relatively short and limited. A new generation of Web searching tools is needed that are designed to support human information behaviors as people seek to resolve their information problems. Improved Web search will come with the design of more effective search tools that support user information behaviors. But also, Web users need to understand the importance of investing time in developing their information behaviors and searching skills to more effectively use complex Web searching tools.

Our research is continuing with large scale data from Alta Vista that we will compare with findings from other Web search engines.

Acknowledgment

We thank Excite.com and Per Gunan Auran of FAST No. in the provision of Web query data sets.

References

Montgomery, A., & Faloutsos, C. (2001). Identifying web browsing trends and patterns. *IEEE Computer* (July, 2001), 94-95.

Silverstein, C., Henzinger, M., Marais, H., & Moricz, M. (1999). Analysis of a Very Large Web Search Engine Query Log. *Association for Computing Machinery, Special Interest Group on Information Retrieval (ACM SIGIR) Forum*, 33, 3.

Spink, A., Batemen, J., & Jansen, B. J. (1999). Searching the Web: Survey of Excite users. *Internet Research: Electronic Networking Applications and Policy*, 9(2), 117-128.

Spink, A., Jansen, B. J., Wolfram, D., & Saracevic, T. From e-sex to e-commerce: Web searching changes. *IEEE Computer*, 35(3), 107-109.

Spink, A., Ozmutlu, H. C., & Ozmutlu, S. (2002). Multitasking information seeking and searching processes. *Journal of the American Society for Information Science and Technology*, 53(8), 639-652.