Association for Information Systems
AIS Electronic Library (AISeL)

AMCIS 2005 Proceedings

Americas Conference on Information Systems (AMCIS)

2005

One Week with a Corporate Search Engine: A Time Based Analysis of Intranet Information Seeking

Dick Stenmark
Goeteborg University, stenmark@informatik.gu.se

One week with a corporate search engine: A time-based analysis of intranet information seeking

Dick Stenmark
Department of Informatics, Göteborg University, Sweden
stenmark@informatik.gu.se

ABSTRACT

In this explorative work, we have focused on understanding information seeking behaviour amongst intranet users. By carrying out a time-based analysis of a week's worth of log data from a corporate-internal search engine, we have been able to observe patterns of usage as it shifts over days and hours. The results show that numbers of started sessions and activities correlate and follow business hours closely but also that the number of terms per query differs significantly over the day but is constant over the week. The number of active users and the number of sessions are higher early in the week and declines as the week progresses, and we also note that frequent search engine users also log more activities per visit. This study shows that intranet seeking behaviour differs from what is known about public web searching. The main contribution is the baseline for more targeted intranet studies that this study provides.

Keywords

Information seeking, intranet search engines, log file analysis.

INTRODUCTION

Even though information retrieval (IR) has undergone a tremendous development over to past decades and has come to include many things beyond mere indexing and retrieval of text, it has still been considered a rather narrow field of interest mainly to librarians and information scientists. The advent of the World Wide Web (hereafter the web) in the early 1990s radically altered this perception (Baeza-Yates and Ribeiro-Neto, 1999). With the web, the IR field was suddenly opened to millions of users with little or no knowledge of traditional search tools (Jansen & Spink, 2003). It was, however, soon discovered that these laymen were not *retrieving* information. Instead these users were *seeking* information, where the latter is a more human-oriented and open-ended process. In information seeking, you do not know whether there exists an answer to your query. The very process of seeking may provide the learning required to satisfy the information need (Choo, Detlor and Turnbull, 2000). Trying to understand how laymen search for information, a body of research on how casual users interact with web search engines is beginning to form. However, as Spink and Jansen (2004) point out; we have only begun to explore this field and more research is needed.

Alongside the development of the web, intranets (i.e., intra-organisational networks based on web technology) have emerged as rapidly growing information environments, and there are commentators that suggest that three out of every four web servers being installed are intended for intranet usage (Gerstner 2002). Obviously, intranet users also need sophisticated tools to help them find information in their organisations, but if little is yet know about information seeking on the web, virtually nothing is known about intranet searching. Our explorative work acknowledges this fact, and in this paper we examine the use of a corporate search engine by carrying out a time-based analysis of a week's worth of search engine log file data.

In this study, we investigate intranet users' information seeking behaviour with respect to different days and different hours of the week. As suggested by Ozmutlu, Spink and Ozmutlu (2004) such data can be useful when trying to optimise web search engine performance by intelligent allocation of resources. The most important aspect of the study is that it furthers our understanding of intranet users by providing details of their search behaviour. The outline of the paper is as follows. In the next section, we account for some related work on search log analysis carried out on public search engine and show that little work has been focusing on understanding search engine usage on intranets. Section three describes the research setting and our methodology. The results from our study are presented in section four, and analysed and discussed in section five. Finally, the paper ends with conclusions in section six.

BACKGROUND AND RELATED WORK

A useful amount of work on how users interact with public search engines has started to compile over the last years. Amongst the first researchers to study this phenomenon at a larger scale were Silverstein, Henzinger, Marais and Moricz (1998) with their analysis of AltaVista usage and Jansen, Spink, Bateman and Saracevic (1998) with their study of Excite. Spink and Jansen claim that although a body of knowledge is starting to emerge, we still know little about how ordinary people search the web. In their words; "[r]esearch on human information behavior on the Web is in its infancy – we have only begun to chart this area of studies" (Spink & Jansen, 2004, p.xii). Through the work of the above and other scholars, we have come to realise that web users differ significantly from librarians and other trained information retrieval (IR) specialists in their search behaviour, and that web-based IR tools therefore need to be studied in their own right.

A related field that may be thought of as a subset of web search is the use of search engines on corporate intranets. Technology-wise, intranets and the public web are identical, and it may therefore be easy, albeit incorrect, to think of intranets as downscaled webs. Stenmark (2005a) argues that, due to cultural differences, intranets differ from the web in many aspects. Before we can extend previous findings on web searching to intranet users, we must know more about this particular user group and their behaviour. Apart from our own work recently started work (cf. Stenmark, 2005b), surprisingly little research has been targeted towards information seeking on intranets.

Hawking, Bailey and Craswell (2000) implemented a text retrieval system in a university intranet. However, their aim was to "reality check" a system that had performed well in lab tests rather than to gain in-depth understanding of intranet search behaviour, and they do not tell us much about the intranet users.

Göker and He (2000) examined search engine logs from Reuter's intranet in order to test methods to detect session boundaries automatically. Their research is highly relevant for all scholars studying search logs, and shall be used in this study as described in the next section, but they did not tell us anything about the intranet users' interactions with the search engine.

Fagin et al. (2003) studied rank aggregation and the effects of different heuristics on the ranking of search results on IBM's intranet by implementing a search engine prototype. They found that intranets are different from the web and that the social processes behind content creation differ between intranet and Internet, but they provided no details on specific intranet search behaviour.

In other words, since intranets are well established and widely used technology that has received little academic interest, there is a need for more in-depth studies of intranet searching.

RESEARCH SETTING AND METHOD

This study is based on search engine log data obtained from the SwedCorp intranet. SwedCorp is a manufacturer of commercial vehicles with some 70,000 employees and offices and factories in many countries around the world. The headquarters are located in Sweden. In 2002, the SwedCorp intranet consisted of more than 1,500 known web servers, and the search engine reported to have indexed over 740,000 documents. Most of these were of HTML pages (approx. 80%) but there were also PDF documents (~15%) and MS-Office documents (~5%).

Since 1998, SwedCorp uses Ultraseek – a commercial search engine originally developed and sold by Infoseek Inc but is now owned by Verity, Inc. Ultraseek does not accept Boolean operators such as AND or OR but instead allows the use of + (plus) and – (minus) to indicate that a term MUST or MUST NOT appear in the document. Quotation marks are used to indicate a string search and all these features may be combined. The Ultraseek log files include the IP address of the user's machine, the date and time of the submission, the query string as entered by the user and some additional parameters not used in this analysis. The server was located in Sweden and all timestamps are in Central European Time (CET), regardless of the user's location. The logs used in our work contained 26,205 activities from 5,644 different IP addresses from midnight October 21st 2002 to midnight October 27th 2002 CET.

Log file analysis has its advantages and its shortcomings. The benefits of using transaction log analysis are that it is unobtrusive and easily accessible. As noted by Hawking et al. (2000), the spontaneous and sporadic characteristics of search behaviour make naturalistic observations of real searchers virtually impossible. It can also be assumed that such monitoring would affect the users' behaviour, if done overtly, or violate their integrity, if done covertly.

On the down side it should be noted that transaction log analysis does not tell us anything about the context in which the searcher operates, and it is therefore difficult to correctly understand the reasons the users have for searching, assess the quality of their results, or the performance of the system (cf. Hawking et al., 2000; Spink et al., 2001). Optimally, transaction log analysis should be complemented by qualitative studies of the users and their environment, and such studies have been

initiated. Nonetheless, transaction log analysis is an established research method (see e.g., Silverstein et al., 1998; Jansen et al., 1998; Spink et al., 2001, Ozmutlu et al., 2004) and to be able to compare and contrast our findings with what is known about public web searchers, we shall adopt a method similar to what has previous been used. That said, it must be acknowledged that this is no easy task since no standardised metrics have been agreed upon and methods and definitions differ between studies (Jansen and Pooch, 2001; Spink et al., 2001).

Some of the more commonly used definitions have also been questioned, most notably the session definition (Göker and He, 2000). Transaction log-based studies often group all activities for one user into one session (e.g., Jansen, Spink and Saracevic, 2000; Jansen and Pooch, 2001; Spink et al., 2001; Jansen and Spink, 2003; Ozmutlu et al., 2004). As pointed out by Göker and He, this results in many unrelated activities being incorrectly grouped together – particularly so when the data spans multiple days. Since we had an entire week's worth of data, the straightforward but naïve session definition used in much of the previous research would have been useless.

Instead, we used the session definition heuristics suggested by Göker and He (2000) to derive automatically more reliable search session. Based on analysis of Reuter's intranet search engine, Göker and He argued that users may engage in several different search sessions during a single day and to separate sessions they suggested that when the idle time between two consecutive user actions exceeds a certain threshold, a session boundary has been detected. Their empirical data suggested that a threshold of 11-15 minutes would generate the most reliable sessions. Having applied their method to our data and calculated the number of sessions generated for an interval threshold of 1-16, 20, 30 and 45 minutes (see figure 1), we noticed that the curve flattened out after 10 minutes, and we therefore chose an idle time threshold of 13 minutes, since this fell in the middle of the interval suggested by Göker and He. This resulted in 11,419 user sessions.

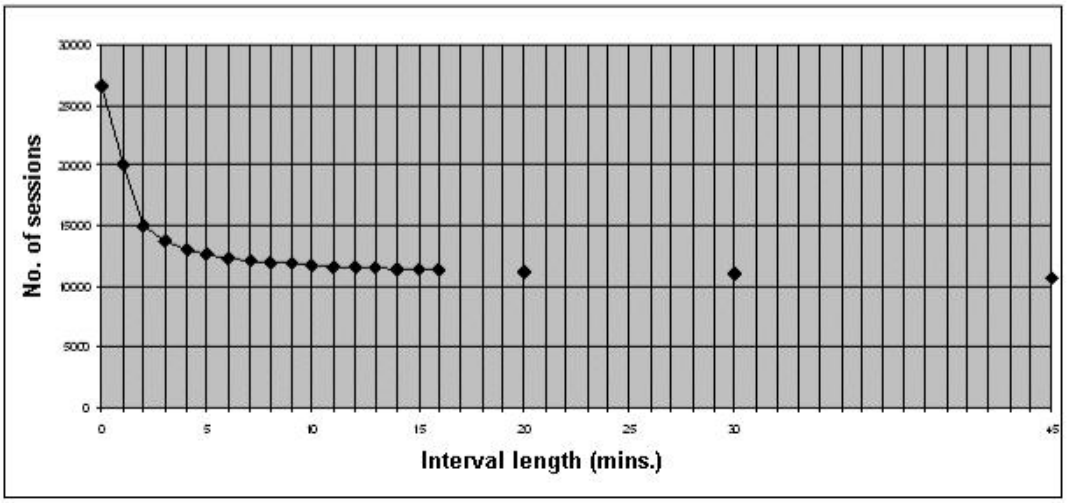

Figure 1: No. of search sessions as a function of inter-session idle interval length

RESULTS

Activities

A search engine user may perform two types of *activities*; submit queries and browse search results. When a query is submitted, the search engine processes the query and displays the first set of results. This generates *one* log entry and we shall refer to it as a query. Usually, there are more results than can be displayed on a single page and the user may access lower ranked results by clicking a Next button (or equivalent). This generates another log entry and we refer to this as a result request. Together, these two actions are referred to as activities.

The number of activities per hour was distributed according to figure 2 (solid line). The activity takes off after 06:00 hours and rises to a peak at 10:00 hours with 2,828 activities. After a dip during lunch hours, the activity is again high between 13:00 and 15:00 hours, before declining in the afternoon. After 18:00 hours, activity is rather low with a minimum of 45 activities at 04:00 hours.

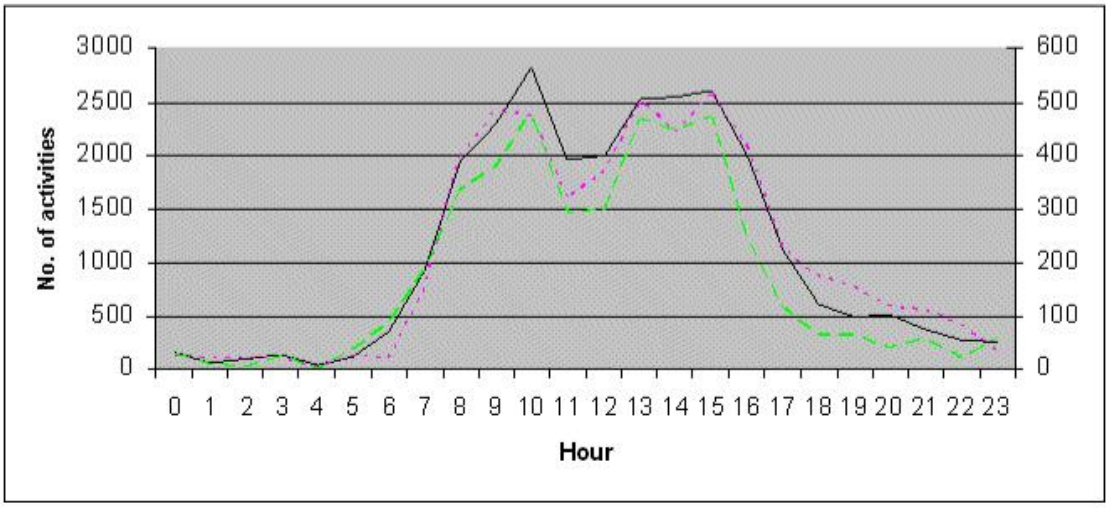

Figure 2: Number of activities per hour for the entire week (solid line/left axis), for Wednesday (dotted line/right axis), and for Friday (dashed line/right axis)

This pattern is repeated for the individual days. At peak hours, approximately 500 activities per hour were recorded (see figure 2, right axis). There are small variations; we see that the afternoon drop occurs somewhat earlier on Friday (dashed line in figure 2) than on the other days.

Figure 3 shows that the distribution of activities per day is rather consistent over the weekdays, with a slight but noticeable drop on Friday. The average number of activities per day Monday-Thursday was 5,428.5 whilst the number of activities on Friday was 4,182; a reduction of 23%. People did not search much during the weekends and the average number of activities during Saturday and Sunday was 154.5.

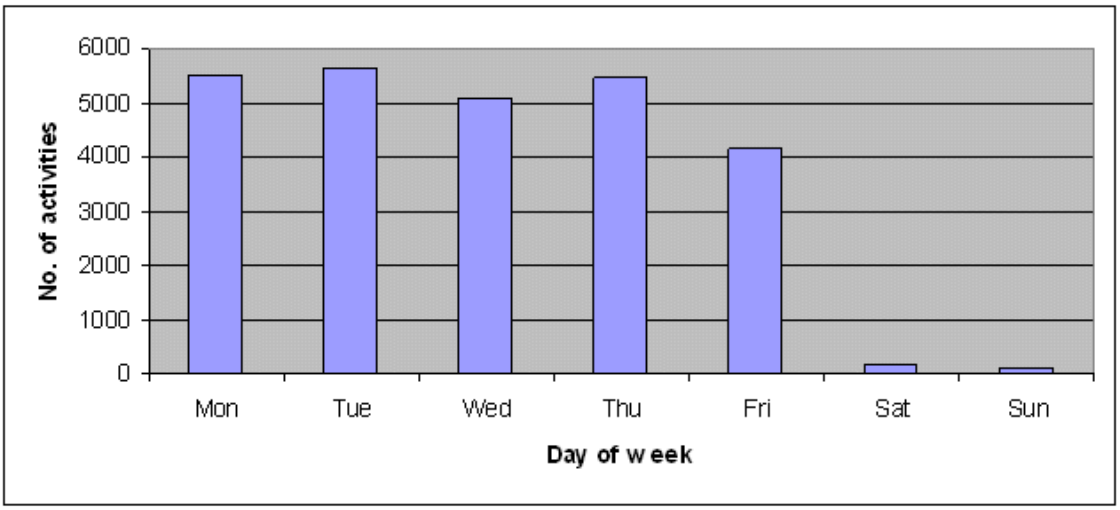

Figure 3: Number of activities per day (Mon-Sun)

Sessions

When examining the number of sessions started per hour (figure 4, dotted line), we noticed that the distribution was almost identical to the activities per hour distribution (figure 4, solid line). Again, the activity was high between 07:00 and 17:00 hours with a noticeable dip around lunchtime. During the peak at 10:00 hours, 1196 sessions were initiated to be compared to the 28 sessions started at 02:00 hours. The afternoon decline followed the distribution seen for activities.

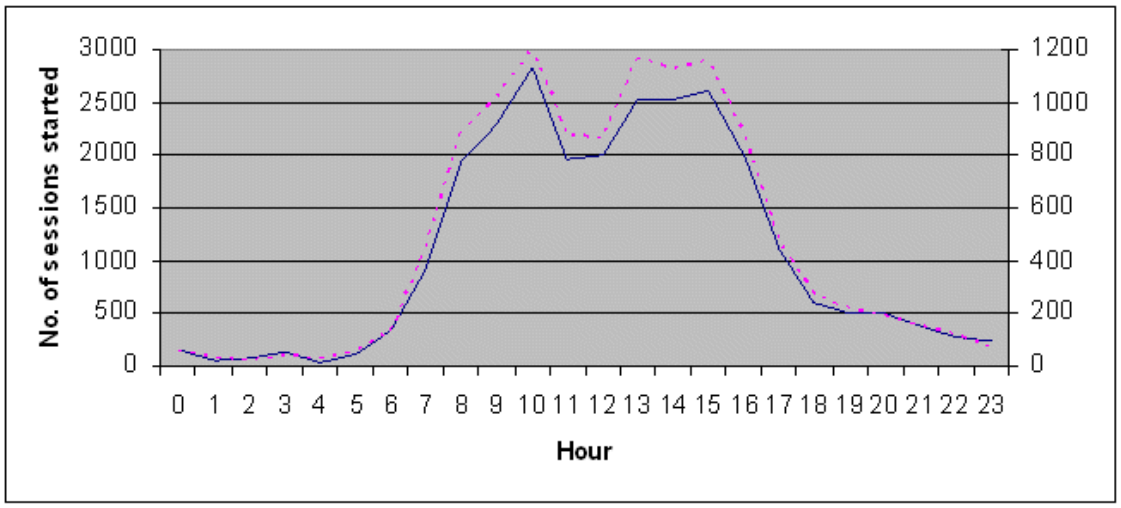

Figure 4: Number of sessions started per hour for the entire week (dotted line/right axis) compared to number of activities per hour (solid line/left axis)

The distribution of sessions started per day (figure 5 below) differed somewhat from the distribution of activities per day (figure 3). The number of sessions started per day had a peak on Monday with 2,511 sessions and thereafter declined steadily thought-out the week. On Friday, 1,875 sessions were started (a drop of 25% compared to Monday) and on Sunday only 52 sessions were initiated. The average number of sessions started Monday-Friday was 2,279.6 and Saturday-Sunday 71.5. The average for the week was 1,642.3.

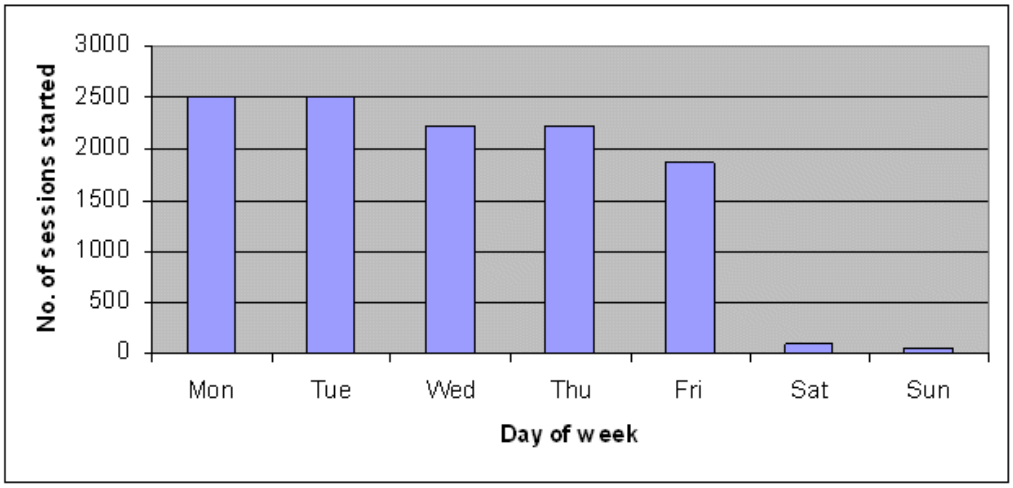

Figure 5: Number of search sessions started per day of the week

Combining figures 3 and 5, the average number of activities per session distributed per day (figure 6) shows that the ratio is pretty consistent, also for the weekend. The average number of activities per session for the entire week is 2.29. Saturday is 14% below and Sunday is 11% above the average.

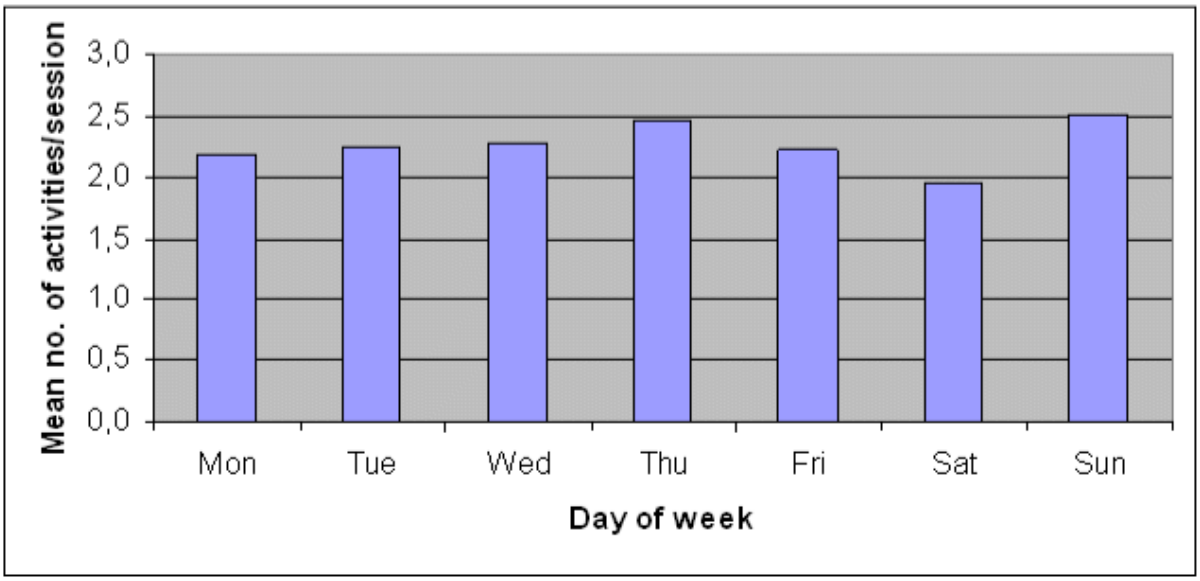

Figure 6: Average number of activities per session distributed over days

Terms

The average number of terms per query was 1.40 but we noticed large variations across time as can be seen in figure 7. The number of terms per query was significantly higher during the night hours. Having removed empty queries (i.e., queries with no search terms at all), we found a peak value of 1.79 terms per query at 01:00 hour. This number declined steadily to a lowest value of 1.26 terms per query at 07:00 hour. From there on, the number of query terms increased slightly throughout the day. With empty queries included, the average obviously drops (dotted line in figure 7) but the distribution remains pretty much the same.

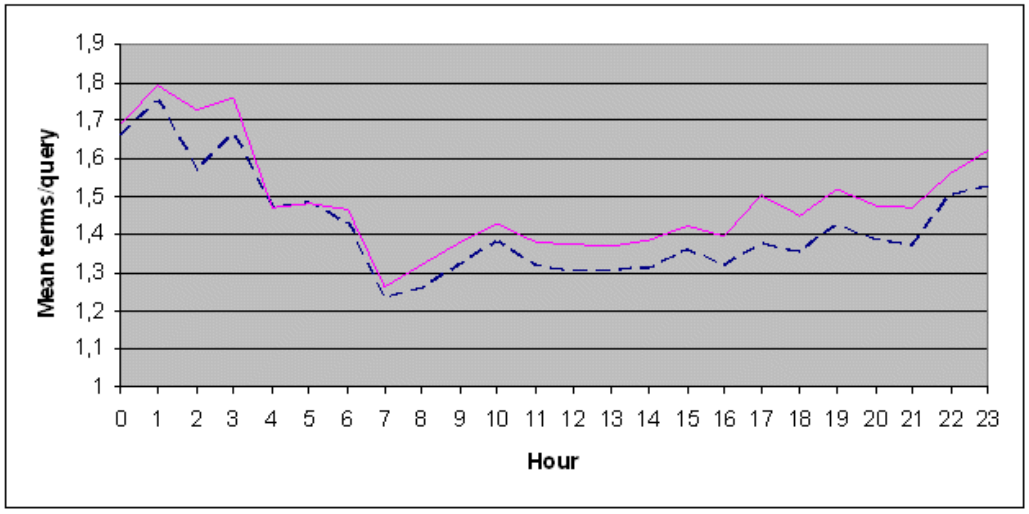

Figure 7: Average number of terms per query displayed per hour
(Solid line= empty queries removed, dotted line= empty queries included)

Looking at the distribution of query length per day (figure 8), we saw that the average query length reached its peak on Friday with 1.43 terms per query. However, the distribution for Monday-Saturday was very consistent, with an average query length of 1.40 and a standard deviation of only 0.016. In contrast, the average query length on Sunday was 1.26 – a good 10% below the other days.

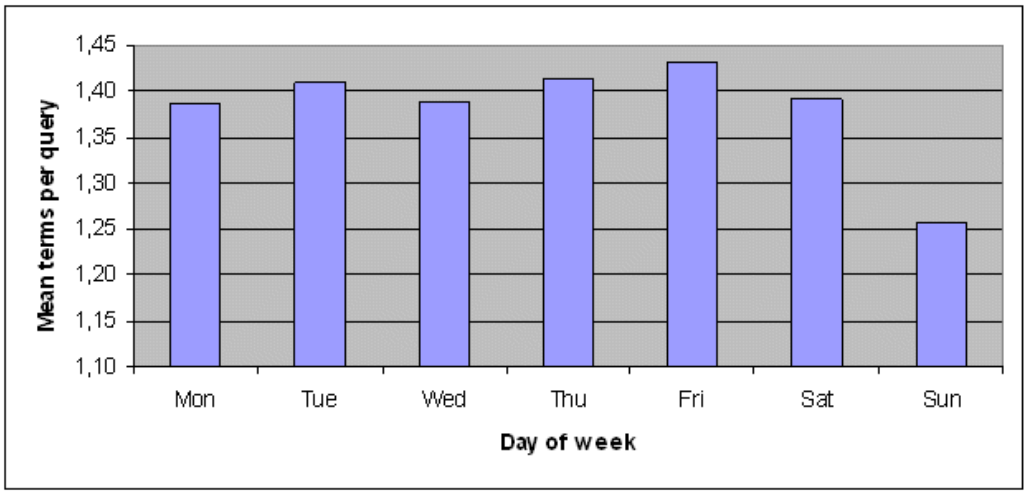

Figure 8: The variation of query length across the week

Users

The majority, or almost 71% of the 5,644 users, were only active during one of the monitored days (figure 9 below). Less than 9% were active on three days or more and only 40 users (0.7%) logged activities during five days. No single user was active on more than five days during the measured week. In 36 of the 40 cases the users were active during Monday through Friday, but four users had patterns including weekend activities.

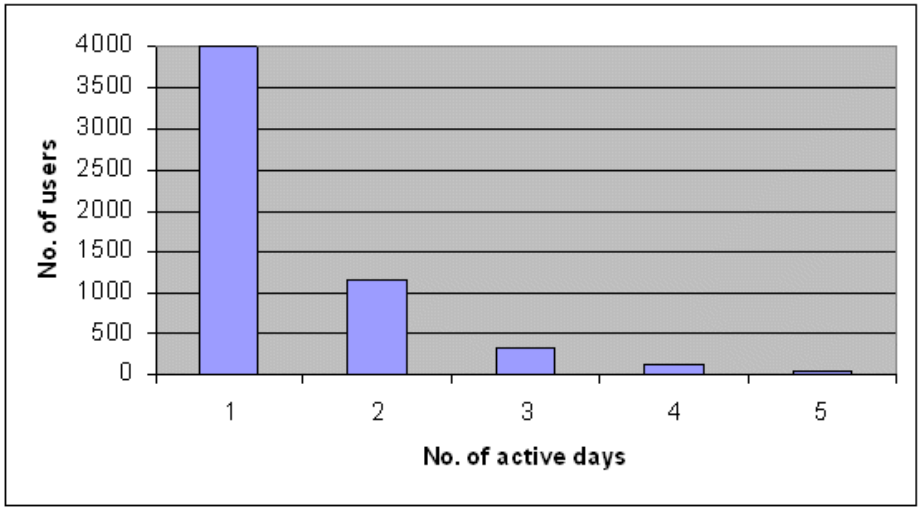

Figure 9: Seventy percent of the users were active only on one day.
Only forty users (0.7%) were active during five days.

Users who were active during five days obviously logged more activities than did those who were active on fewer days. What is more noteworthy is that the number of activities per day also differed. As can be seen from figure 10 (solid line), the more days the users had been active, the more activities per day they logged. This is true also for the number of session (figure 10, dotted line).

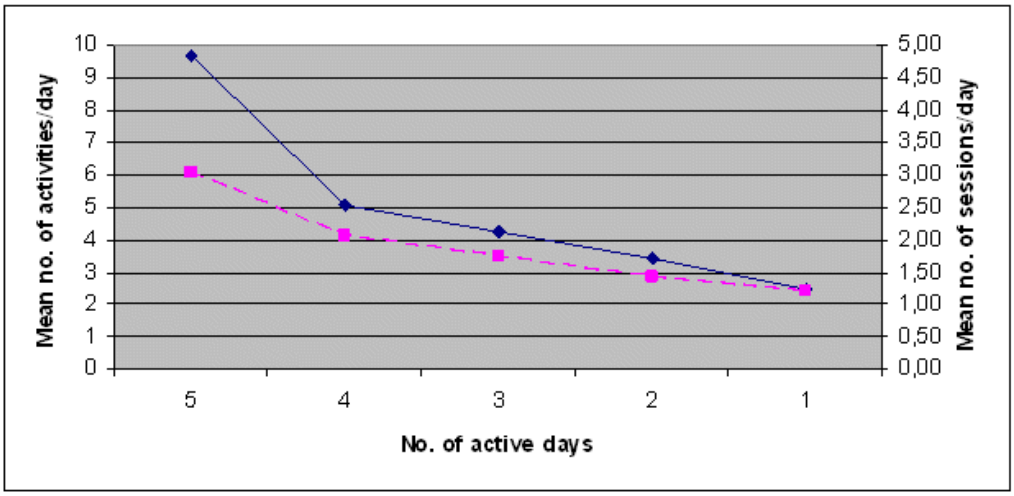

Figure 10: Average number of activities (solid line) and sessions (dotted line) per day as a function of number of active days during the week

When analysing how many users were active during the individual days, we found the pattern illustrated in figure 11. As with activities and sessions, we noticed that the number of active users was low during the weekend with only 68 and 44 users accessing the search engine during Saturday and Sunday, respectively. Noticeable was also the relatively low number of users on Friday. The average number of active users per day for Monday-Thursday was 1652, whilst only 1304 users logged activities during Friday; a decline of 21%.

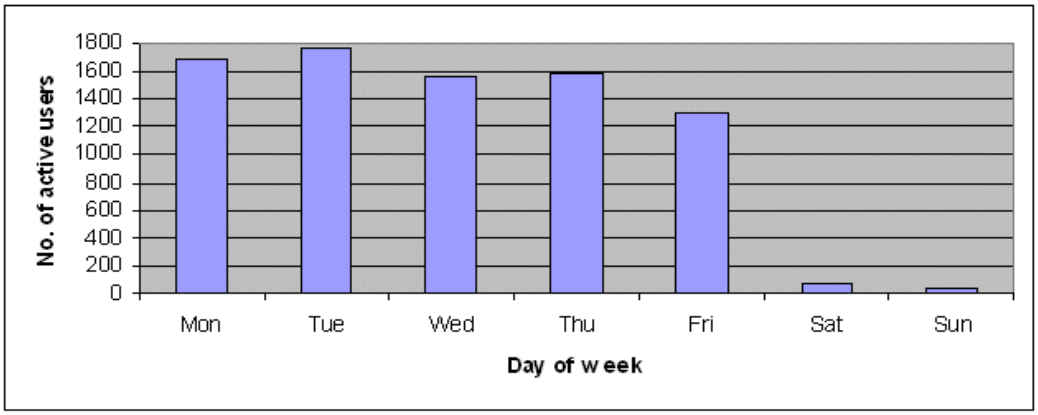

Figure 11: Number of active users distributed over days

The results per day and hour are summarised in tables 1 and 2. Number of activities and started session were significantly lower during the weekend whereas the average query length was more evenly distributed (table 1).

Table 1: A summary of the results as distributed per day. Values deviating from norm are marked

Day of week	*No. of daily session arrivals*	*No. of daily activities*	*Mean terms per query*
Monday	2511	5505	1.386158
Tuesday	2503	5639	1.410481
Wednesday	2234	5091	1.389019
Thursday	2230	5479	1.414207
Friday	1875	4182	1.431134
Saturday	91	178	1.391892
Sunday	52	131	1.257143

Number of started sessions and number of activities both dropped during lunch hours whilst query length had its extreme values way outside this interval (table 2).

Table 2: A summary of the results as distributed per hour. Values deviating from norm are marked.

Hour of day	*No. of hourly session arrivals*	*No. of hourly activities*	*Mean terms per query*
00:00-01:00	60	159	1.688679
01:00-02:00	30	51	1.790698
02:00-03:00	28	89	1.728814
03:00-04:00	38	136	1.758621
04:00-05:00	29	45	1.472222
05:00-06:00	56	123	1.480392
06:00-07:00	130	349	1.464286
07:00-08:00	440	918	1.264286
08:00-09:00	895	1936	1.321877
09:00-10:00	1011	2292	1.380541
10:00-11:00	1196	2828	1.428002
11:00-12:00	879	1957	1.379531
12:00-13:00	869	2001	1.376793
13:00-14:00	1165	2536	1.369533
14:00-15:00	1123	2542	1.386510
15:00-16:00	1158	2609	1.422211
16:00-17:00	871	1996	1.398072
17:00-18:00	468	1122	1.505882
18:00-19:00	276	606	1.448473
19:00-20:00	220	497	1.518414
20:00-21:00	197	507	1.475410
21:00-22:00	156	377	1.470175
22:00-23:00	125	280	1.560976
23:00-24:00	76	249	1.623377

DISCUSSION

In this study, we have analysed one week's worth of log data from a corporate intranet search engine to find patterns of use and to compare these to those reported in studies of public search engine usage. Due to the enormous amount of log data generated by public search engines, studies of public search engines have been forced to use data from a single day or less. On corporate intranets, the amount of data is more manageable which has enabled us to carry out this time-based analysis. We are aware of the fact that our study differs methodologically form others and that this makes comparisons of data more challenging but where possible we shall try to relate our findings to the time-based analysis of Excite and FAST carried out by Ozmutlu et al. (2004).

Activities

The increase in activities after 06:00 hours that we recorded is analogous to what Ozmutlu et al. (2004) observed for users of the FAST search engine. Ozmutlu et al's conclusion that mid-morning to noon seem to be the prime-time for web searching seems to hold also for intranet searchers. However, whereas the activities of their Internet users showed a slow and steady decline throughout the day, our data shows a much more dramatic drop in the afternoon. A possible explanation is that Internet users can access search engines also from their homes while intranet users may have limited access when leaving the office. This explanation would also explain the low activity level during the weekend. Another possible explanation is intranet information typically is work related and thus not needed during off-hours. There may be other explanations as well and more research on intranet searching is required to understand fully these differences.

Sessions

Ozmutlu et al. (2004) found session arrivals and query arrivals to correlate. This holds also for our data, although our pattern is completely different. Analysing the FAST logs, Ozmutlu et al. found session and query arrivals per hour to peak at around 9 am and thereafter decrease steadily until 7 am the following day, when the numbers climb rapidly. Our definition of a

session is obviously very different from the one used by Ozmutlu et al., and it is likely that this has had an impact on the result. Exactly how this has affected the outcome is difficult to say. Our data shows a later peak hour, an obvious lunch drop, and a much more distinct drop in the afternoon. The morning rise and the afternoon drop coincide with regular office hours, again suggesting that intranet searching is not carried out from outside the office.

A novel finding is that the number of started sessions decreases throughout the week. From its highest value of 2,511 sessions started on Monday it drops steadily to a low of 1,875 on Friday. Judging from our data, it appears users are more search prone early in the week.

Terms

Ozmutlu et al. concluded that the number of terms per query was very consistent across hours. We have a very different result. The average query length measured at 01:00 hours is 42% longer than the average query length measure at 07:00 hours. We have yet no explanation for this in our opinion surprising result and further investigations are required to find out whether these multiple-term queries were submitted by European office workers staying up late, European night shift workers or corporate members located on other continents. 01:00 hours Central European Time translates to afternoon/evening in the U.S. and early morning in Asia. However, we see that the average query term length is consistent across days – Sundays excepted.

The Swedish language uses compound words which means that queries such “torque wrench” or “digital tape library” which in English require two or three terms, would be expressed using only one Swedish word. If we assume the portion of non-Swedes to be higher at 01:00 than during daytime, this could explain the fluctuation in terms per query. A language-based analysis of the query terms may be useful even though many Swedes use English terms. An interesting future study would be to isolate English and non-English speaking intranet users and compare their use of search terms.

Users

Only one out of every four users accessed the search engine more than once during the measured week. This is new and interesting knowledge since previous studies of public search engines have used data from (portion of) a single day and thus not been able to perform this sort of (long) time-based analysis.

On average, these infrequent users engaged in 1.2 sessions and 2.47 activities during their single day of searching. In sharp contrast, the most frequent searchers, who accessed the search engine five times a week, on average engaged in 3.03 sessions and 9.7 activities per day. Our data suggest that the more often a user interacts with the search engine, the more active he or she is during those interactions; in terms of number of sessions, number of activities, and number of activities per session. An interesting, and pending, question is whether the high-activity users search more because they are good or because they are so bad at searching. A closer examination of the users may reveal different categories of searchers that may allow tailoring of search tools for particular groups of users.

Implications and limitations

The contribution of this pioneering work is that it offers a baseline for future studies of intranet information seeking and search engine usage. We have tried to relate out work to studies of public search engine use, since no studies of intranet search engines have been published, but other researchers of intranets can now compare and relate their findings to those reported in this paper and thereby further our accumulated understanding. Our studied of web search engines in *one* intranet is not sufficient to draw general conclusions, since usage may differ significantly between intranets. We therefore invite more studies of intranet search engine usage and hope that this work can inspire others to engage in similar studies.

As mentioned above, log file analysis tells us nothing about the context in which the searcher operates, nor does it explain *why* we see certain patterns. In the above discussion, we have pointed to a number of unanswered questions and one may argue that our research has raised more questions than it has answered. This, we believe, is unavoidable, since intranet searching represents an uncharted territory and explorative research is needed to understand the issues we are facing. In a parallel study, we are deploying qualitative studies to complement the transaction log analysis used in this work. Hopefully such triangulation will further our understanding of web searching in general and intranet searching in particular.

Another limitation is the difference in methodology between our study and previous work. Not only did we define sessions differently, but we also did not have in our logs some parameters reported by other researchers. The fact that there are no standardised metrics, methods, or definitions is a recognised and (to some extent) debated problem in studies of information seeking (cf. Jansen and Pooch, 2001; Spink et al., 2001). We hope that our work has contributed to this discussion, and that, as the body of knowledge continues to grow, a consensus on these matters will emerge.

CONCLUSIONS

Surprisingly little academic attention has been paid to the fact that intranet users also need sophisticated tools to help them find information in their organisations. Without knowing how these users interact with their internal search engines, we do not know what features to strengthen and what to remove. In this explorative work, we have examined the use of a corporate search engine by doing a time-based analysis of a week's worth of search engine log file data.

We have found that search sessions and search activities follow the office hours closely; a sharp increase around 7 am, a dip around lunchtime, and a sharp drop at 5 pm. These patterns differ from what has previously been reported for web searchers.

Whereas the number of query terms used in web searches is consistent across the day, intranet queries vary significantly in length. From seven o'clock in the morning, where we have the shortest average query length, the number of query terms increases slowly until its peak at 1 am at night.

Usage – in terms of both active users and number of started sessions – is higher earlier in the week and declines as the week progresses. Since few studies have examined more than a day's worth of log data, these results have not previously been reported.

Most intranet users access the search engine only once a week. However, there are also users who access the search engine several times daily. Even more interestingly, we found that the more days per week a user interacts with the search engine, the more sessions and activities per day they log.

ACKNOWLEDGMENTS

This work was funded by the Swedish Council for Work and Social Research via grant #004-1268. The author is grateful for all valuable comments received during the review process.

REFERENCES

1. Baeza-Yates, R. and Ribiero-Neto, B. (1999) Modern Information Retrieval, Addison-Wesley.
2. Choo, C. W., Detlor, B. and Turnbull, D. (2000) Web Work: Information seeking and knowledge work on the World Wide Web. Kluwer Academic Publisher.
3. Fagin, R., Kumar, R., McCurley, K., Novak, J., Sivakumar, D., Tomlin, J. and Williamson, D. (2003) Searching the Corporate Web, *Proceedings of WWW2003*, Budapest, Hungary, 366-375.
4. Gerstner, J. (2002) Intranets mean Business, *Communication World*, 19, 2, 14-17.
5. Göker, A. and He, D. (2000) Analysing Web Search Logs to Determine Session Boundaries for User-Oriented Learning. *Proceedings of Adaptive Hypermedia and Adaptive Web-based Systems*, Trento, Italy, 319-322.
6. Hawking, D., Bailey, P. and Craswell, N. (2000) An intranet reality check for TREC ad hoc, Technical report: CSIRO Mathematical and Information Sciences.
7. Jansen, B. and Pooch, U. (2001) A review of web searching studies and a framework for future research, *Journal of the American Society for Information Science*, 52, 3, 235-246.
8. Jansen, B. and Spink, A. (2003) An Analysis of Web Documents Retrieved and Viewed. *Proceedings of ICIC'03*, Las Vegas, NE, 65-69.
9. Jansen, B.J., Spink, A., Bateman, J. and Saracevic, T. (1998) Real life information retrieval: A study of user queries on the web. *ACM SIGIR Forum*, 32, 1, 5-17.
10. Jansen, B., 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 and management*, 36, 207-227.
11. Ozmutlu, S., Spink, A. and Ozmutlu, H. (2004) A day in the life of Web searching: an exploratory study. *Information processing and management*, 40, 319-345.
12. Silverstein C., Henzinger, M., Marais, H. and Moricz, M. (1998) Analysis of a Very Large AltaVista Query Log, *Digital SRC Technical Note #1998-014*, October 26.
13. Spink, A. and Jansen, B. (2004) Web Search: Public searching of the web. Kluwer Academic Publisher.
14. Spink, A., Wolfram, D., Jansen, B. and Saracevic, T. (2001) Searching the web: The public and their queries. *Journal of the American Society for Information Science*, 52, 3, 226-234.
15. Stenmark, D. (2005a) How intranets differ from the web: organisational culture's effect on technology. *Proceedings of ECIS 2005*, Regensburg, Germany, May 26-28 2005.
16. Stenmark, D. (2005b) Searching the intranet: Corporate users and their queries. To appear in *Proceedings of ASIS&T 2005*, Charlotte, North Carolina Oct. 28 – Nov. 2 2005.