

Investigating the Querying and Browsing Behavior of Advanced Search Engine Users

Ryen W. White
Microsoft Research
One Microsoft Way
Redmond, WA 98052
ryenw@microsoft.com

Dan Morris
Microsoft Research
One Microsoft Way
Redmond, WA 98052
dan@microsoft.com

ABSTRACT

One way to help all users of commercial Web search engines be more successful in their searches is to better understand what those users with greater search expertise are doing, and use this knowledge to benefit everyone. In this paper we study the interaction logs of advanced search engine users (and those not so advanced) to better understand how these user groups search. The results show that there are marked differences in the queries, result clicks, post-query browsing, and search success of users we classify as advanced (based on their use of query operators), relative to those classified as non-advanced. Our findings have implications for how advanced users should be supported during their searches, and how their interactions could be used to help searchers of all experience levels find more relevant information and learn improved searching strategies.

Categories and Subject Descriptors

H.3.3 [Information Search and Retrieval]: query formulation, search process, relevance feedback.

General Terms

Experimentation, Human Factors.

Keywords

Query syntax, advanced search features, expert searching.

1. INTRODUCTION

The formulation of query statements that capture both the salient aspects of information needs and are meaningful to Information Retrieval (IR) systems poses a challenge for many searchers [3]. Commercial Web search engines such as Google, Yahoo!, and Windows Live Search offer users the ability to improve the quality of their queries using query operators such as quotation marks, plus and minus signs, and modifiers that restrict the search to a particular site or type of file. These techniques can be useful in improving result precision yet, other than via log analyses (e.g., [15][27]), they have generally been overlooked by the research community in attempts to improve the quality of search results.

IR research has generally focused on alternative ways for users to specify their needs rather than increasing the uptake of advanced syntax. Research on practical techniques to supplement existing

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGIR'07, July 23–27, 2007, Amsterdam, The Netherlands.

Copyright 2007 ACM 978-1-59593-597-7/07/0007...\$5.00.

search technology and support users has been intensifying in recent years (e.g. [18][34]). However, it is challenging to implement such techniques at large scale with tolerable latencies.

Typical queries submitted to Web search engines take the form of a series of tokens separated by spaces. There is generally an implied Boolean AND operator between tokens that restricts search results to documents containing all query terms. De Lima and Pedersen [7] investigated the effect of parsing, phrase recognition, and expansion on Web search queries. They showed that the automatic recognition of phrases in queries can improve result precision in Web search. However, the value of advanced syntax for typical searchers has generally been limited, since most users do not know about advanced syntax or do not understand how to use it [15]. Since it appears operators can help retrieve relevant documents, further investigation of their use is warranted.

In this paper we explore the use of query operators in more detail and propose alternative applications that do not require all users to use advanced syntax explicitly. We hypothesize that searchers who use advanced query syntax demonstrate a degree of search expertise that the majority of the user population does not; an assertion supported by previous research [13]. Studying the behavior of these advanced search engine users may yield important insights about searching and result browsing from which others may benefit.

Using logs gathered from a large number of consenting users, we investigate differences between the search behavior of those that use advanced syntax and those that do not, and differences in the information those users target. We are interested in answering three research questions:

- (i) Is there a relationship between the use of advanced syntax and other characteristics of a search?
- (ii) Is there a relationship between the use of advanced syntax and post-query navigation behaviors?
- (iii) Is there a relationship between the use of advanced syntax and measures of search success?

Through an experimental study and analysis, we offer potential answers for each of these questions. A relationship between the use of advanced syntax and any of these features could support the design of systems tailored to advanced search engine users, or use advanced users' interactions to help non-advanced users be more successful in their searches.

We describe related work in Section 2, the data we used in this log-based study in Section 3, the search characteristics on which we focus our analysis in Section 4, and the findings of this analysis in Section 5. In Section 6 we discuss the implications of this research, and we conclude in Section 7.

2. RELATED WORK

Factors such as lack of domain knowledge, poor understanding of the document collection being searched, and a poorly developed information need can all influence the quality of the queries that users submit to IR systems ([24],[28]). There has been a variety of research into different ways of helping users specify their information needs more effectively. Belkin et al. [4] experimented with providing additional space for users to type a more verbose description of their information needs. A similar approach was attempted by Kelly et al. [18], who used clarification forms to elicit additional information about the search context from users. These approaches have been shown to be effective in best-match retrieval systems where longer queries generally lead to more relevant search results [4]. However, in Web search, where many of the systems are based on an extended Boolean retrieval model, longer queries may actually hurt retrieval performance, leading to a small number of potentially irrelevant results being retrieved. It is not simply sufficient to request more information from users; this information must be of better *quality*.

Relevance Feedback (RF) [22] and interactive query expansion [9] are popular techniques that have been used to improve the quality of information that users provide to IR systems regarding their information needs. In the case of RF, the user presents the system with examples of relevant information that are then used to formulate an improved query or retrieve a new set of documents. It has proven difficult to get users to use RF in the Web domain due to difficulty in conveying the meaning and the benefit of RF to typical users [17]. Query suggestions offered based on query logs have the potential to improve retrieval performance with limited user burden. This approach is limited to re-executing popular queries, and searchers often ignore the suggestions presented to them [1]. In addition, both of these techniques do not help users *learn* to produce more effective queries.

Most commercial search engines provide advanced query syntax that allows users to specify their information needs in more detail. Query modifiers such as '+' (plus), '-' (minus), and '""' (double quotes) can be used to emphasize, deemphasize, and group query terms. Boolean operators (AND, OR, and NOT) can join terms and phrases, and modifiers such as "site:" and "link:" can be used to restrict the search space. Queries created with these techniques can be powerful. However, this functionality is often hidden from the immediate view of the searcher, and unless she knows the syntax, she must use text fields, pull-down menus and combo boxes available via a dedicated "advanced search" interface to access these features.

Log-based analysis of users' interactions with the Excite and AltaVista search engines has shown that only 10-20% of queries contained any advanced syntax [14][25]. This analysis can be a useful way of capturing characteristics of users interacting with IR systems. Research in user modeling [6] and personalization [30] has shown that gathering more information about users can improve the effectiveness of searches, but require more information about users than is typically available from interaction logs alone. Unless coupled with a qualitative technique, such as a post-session questionnaire [23], it can be difficult to associate interactions with user characteristics. In our study we conjecture that given the difficulty in locating advanced search features within the typical search interface, and the potential problems in understanding the syntax, those users that do use advanced syntax regularly represent a distinct class of searchers who will exhibit other common search behaviors.

Other studies of advanced searchers' search behaviors have attempted to better understand the strategic knowledge they have acquired. However, such studies are generally limited in size (e.g., [13][19]) or focus on domain expertise in areas such as healthcare or e-commerce (e.g., [5]). Nonetheless, they can give valuable insight about the behaviors of users with domain, system, or search expertise that exceeds that of the average user. Querying behavior in particular has been studied extensively to better understand users [31] and support other users [16].

In this paper we study other search characteristics of users of advanced syntax in an attempt to determine whether there is anything different about how these search engine users search, and whether their searches can be used to benefit those who do not make use of the advanced features of search engines. To do this we use interaction logs gathered from large set of consenting users over a prolonged period.

In the next section we describe the data we use to study the behavior of the users who use advanced syntax, relative to those that do not use this syntax.

3. DATA

To perform this study we required a description of the querying and browsing behavior of many searchers, preferably over a period of time to allow patterns in user behavior to be analyzed. To obtain these data we mined the interaction logs of consenting Web users over a period of 13 weeks, from January to April 2006. When downloading a partner client-side application, the users were invited to consent to their interaction with Web pages being anonymously recorded (with a unique identifier assigned to each user) and used to improve the performance of future systems.¹ The information contained in these log entries included a unique identifier for the user, a timestamp for each page view, a unique browser window identifier (to resolve ambiguities in determining which browser a page was viewed), and the URL of the Web page visited. This provided us with sufficient data on querying behavior (from interaction with search engines), and browsing behavior (from interaction with the pages that follow a search) to more broadly investigate search behavior.

In addition to the data gathered during the course of this study we also had relevance judgments of documents that users examined for 10,680 unique query statements present in the interaction logs. These judgments were assigned on a six-point scale by trained human judges at the time the data were collected. We use these judgments in this analysis to assess the relevance of sites users visited on their browse trail away from search result pages.

We studied the interaction logs of 586,029 unique users, who submitted millions of queries to three popular search engines – Google, Yahoo!, and MSN Search – over the 13-week duration of the study. To limit the effect of search engine bias, we used four operators common to all three search engines: + (plus), - (minus), "" (double quotes), and "site:" (to restrict the search to a domain or Web page) as advanced syntax. 1.12% of the queries submitted contained at least one of these four operators. 51,080 (8.72%) of users used query operators in any of their queries. In the remainder of this paper, we will refer to these users as "advanced" searchers. We acknowledge that the direct relationship between query syntax usage and search expertise has only been studied

¹ It is worth noting that if users did not provide their consent, then their interaction was not recorded and analyzed in this study.

(and shown) in a few studies (e.g., [13]), but we feel that this is a reasonable criterion for a log-based investigation. We conjecture that these “advanced” searchers do possess a high level of search expertise, and will show later in the paper that they demonstrate behavioral characteristics consistent with search expertise.

To handle potential outlier users that may skew our data analysis, we removed users who submitted fewer than 50 queries in the study’s 13-week duration. This left us with 188,405 users – 37,795 (20.1%) advanced users and 150,610 (79.9%) non-advanced users – whose interactions we study in more detail. If significant differences emerge between these groups, it is conceivable that these interactions could be used to automatically classify users and adjust a search system’s interface and result weighting to better match the current user.

The privacy of our volunteers was maintained throughout the entire course of the study: no personal information was elicited about them, participants were assigned a unique anonymous identifier that could not be traced back to them, and we made no attempt to identify a particular user or study individual behavior in any way. All findings were aggregated over multiple users, and no information other than consent for logging was elicited.

To find out more about these users we studied whether those using advanced syntax exhibited other search behaviors that were not observed in those who did not use this syntax. We focused on querying, navigation, and overall search success to compare the user groups. In the next section we describe in more detail the search features that we used.

4. SEARCH FEATURES

We elected to choose features that described a variety of aspects of the search process: queries, result clicks, post-query browsing, and search success. The query and result-click characteristics we chose to examine are described in more detail in Table 1.

Table 1. Query and result-click features (per user).

| Feature | Meaning |
|-----------------------------|--|
| Queries Per Second (QPS) | Avg. number of queries per second between initial query and end-of-session |
| Query Repeat Rate (QRR) | Fraction of queries that are repeats |
| Query Word Length (QWL) | Avg. number of words in query |
| Queries Per Day (QPD) | Avg. number of queries per day |
| Avg. Click Position (ACP) | Avg. rank of clicked results |
| Click Probability (CP) | Ratio of result clicks to queries |
| Avg. Seconds To Click (ASC) | Avg. search to result click interval |

These seven features give us a useful overview of users’ direct interactions with search engines, but not of how users are looking for relevant information beyond the result page or how successful they are in locating relevant information. Therefore, in addition to these characteristics we also studied some relevant aspects of users’ post-query browsing behavior. To do this, we extracted search *trails* from the interaction logs described in the previous section. A search trail is a series of visited Web pages connected via a hyperlink trail, initiated with a search result page and terminating on one of the following events: navigation to any page

not linked from the current page, closing of the active browser window, or a session inactivity timeout of 30 minutes. More detail on the extraction of the search trails are provided in previous work [33]. In total, around 12.5 million search trails (containing around 60 million documents) were extracted from the logs for all users. The median number of search trails per user was 30. The median number of steps in the trails was 3. All search trails contained one search result page and at least one page on a hyperlink trail leading from the result page.

The extraction of these trails allowed us to study aspects of post-query browsing behavior, namely the average duration of users’ search sessions, the average duration of users’ search trails, the average display time of each document, the average number of steps in users’ search trails, the number of branches in users’ navigation patterns, and the number of “back” operations in users’ search trails. All search trails contain at least one “branch” representing any forward motion on the browse path. A trail can have additional branches if the user clicks the browser’s “back” button and immediately proceeds forward to another page prior to the next (if any) back operation. The post-query browsing features are described further in Table 2.

Table 2. Post-query browsing features (per trail).

| Feature | Meaning |
|----------------------|--|
| Session Seconds (SS) | Average session length (in seconds) |
| Trail Seconds (TS) | Average trail length (in seconds) |
| Display Seconds (DS) | Average display time for each page on the trail (in seconds) |
| Num. Steps (NS) | Average number of steps from the page following the results page to the end of the trail |
| Num. Branches (NB) | Average number of branches |
| Num. Backs (NBA) | Average number of “back” operations |

As well as using these attributes of users’ interactions, we also used the relevance judgments described earlier in the paper to measure the degree of search success based on the relevance judgments assigned to pages that lie on the search trail. Given that we did not have access to relevance assessments from our users, we approximated these assessments using judgments collected as part of ongoing research into search engine performance.² These judgments were created by trained human assessors for 10,680 unique queries. Of the 1,420,625 steps on search trails that started with any one of these queries, we have relevance judgments for 802,160 (56.4%). We use these judgments to approximate search success for a given trail in a number of ways. In Table 3 we list these measures.

² Our assessment of search success is fairly crude compared to what would have been possible if we had been able to contact our subjects. We address this problem in a manner similar to that used by the Text Retrieval Conference (TREC) [21], in that since we cannot determine perceived search success, we approximate search success based on assigned relevance scores of visited documents.

Table 3. Relevance judgment measures (per trail).

| Measure | Meaning |
|---------|--|
| First | Judgment assigned to the first page in the trail |
| Last | Judgment assigned to the last page in the trail |
| Average | Average judgment across all pages in the trail |
| Maximum | Maximum judgment across all pages in the trail |

These measures are used during our analysis to estimate the relevance of the pages viewed at different stages in the trails, and allow us to estimate search success in different ways. We chose multiple measures, as users may encounter relevant information in many ways and at different points in the trail (e.g., single highly-relevant document or gradually over the course of the trail).

The features described in this section allowed us to analyze important attributes of the search process that must be better understood if we are to support users in their searching. In the next section we present the findings of the analysis.

5. FINDINGS

Our analysis is divided into three parts: analysis of query behavior and interaction with the results page, analysis of post-query navigation behavior, and search success in terms of locating judged-relevant documents. Parametric statistical testing is used, and the level of significance for the statistical tests is set to .05.

5.1 Query and result-click behavior

We were interested in comparing the query and result-click behaviors of our advanced and non-advanced users. In Table 4 we show the mean average values for each of the seven search features for our users. We use p_{advanced} to denote the percentage of all queries from each user that contains advanced syntax (i.e., $p_{\text{advanced}} = 0\%$ means a user never used advanced syntax). The table shows values for users that do not use query operators (0%), users who submitted at least one query with operators ($\geq 0\%$), through to users whose queries contained operators at least three-quarters of the time ($\geq 75\%$).

Table 4. Query and result click features (per user).

| Feature | p_{advanced} | | | | |
|---------|-----------------------|--------------|-------------|-------------|-------------|
| | 0% | > 0% | $\geq 25\%$ | $\geq 50\%$ | $\geq 75\%$ |
| QPS | .028 | .010 | .012 | .013 | .015 |
| QRR | .53 | .57 | .58 | .61 | .62 |
| QWL | 2.02 | 2.83 | 3.40 | 3.66 | 4.04 |
| QPD | 2.01 | 3.52 | 2.70 | 2.66 | 2.31 |
| ACP | 6.83 | 9.12 | 10.09 | 10.17 | 11.37 |
| CP | .57 | .51 | .47 | .47 | .47 |
| ASC | 87.71 | 88.16 | 112.44 | 102.12 | 79.13 |
| %Users | 79.90% | 20.10% | .79% | .18% | .04% |

We compared the query and result click features of users who did not use any advanced syntax ($p_{\text{advanced}} = 0\%$) in any of their queries with those who used advanced syntax in at least one query ($p_{\text{advanced}} > 0\%$). The columns corresponding to these two groups are bolded in Table 4. We performed an independent measures t-test between these groups for each of the features. Since this

analysis involved many features, we use a Bonferroni correction to control the experiment-wise error rate and set the *alpha level* (α) to .007, i.e., .05 divided by the number of features. This correction reduces the number of Type I errors i.e., rejecting null hypotheses that are true. All differences between the groups were statistically significant (all $t(188403) \geq 2.81$, all $p \leq .002$). However, given the large sample sizes, all differences in the means were likely to be statistically significant. We applied a Cohen's d -test to determine the effect size for each of the comparisons between the advanced and non-advanced user groups. Ordering in descending order by effect size, the main findings are that relative to non-advanced users, advanced search engine users:

- Query less frequently in a session ($d = 1.98$)
- Compose longer queries ($d = .69$)
- Click further down the result list ($d = .67$)
- Submit more queries per day ($d = .49$)
- Are less likely to click on a result ($d = .32$)
- Repeat queries more often ($d = .16$)

The increased likelihood that advanced search engine users will click further down the result list implies that they may be less trusting of the search engines' ability to rank the most relevant document first, that they are more willing to explore beyond the most popular pages for a given query, that they may be submitting different types of queries (e.g., informational rather than navigational), or that they may have customized their search settings to display more than only the default top-10 results. Many of the findings listed are consistent with those identified in other studies of advanced searchers' querying and result-click behaviors [13][34]. Given that the only criteria we employed to classify a user as an advanced searcher was their use of advanced syntax, it is certainly promising that this criterion seems to identify users that interact in a way consistent with that reported previously for those with more search expertise.

As mentioned earlier, the advanced search engine users for which the average values shown in Table 4 are computed are those who submit 50 or more queries in the 13 week duration of the data collection *and* submit at least one query containing advanced query operators. In other words, we consider users whose percentage of queries containing advanced syntax, p_{advanced} , is greater than zero. The use of query operators in any queries, regardless of frequency, suggests that a user knows about the existence of the operators, and implies a greater degree of familiarity with the search system. We further hypothesized that users whose queries *more* frequently contained advanced syntax may be *more* advanced search engine users. To test this we investigated varying the query threshold required to qualify for advanced status (p_{advanced}). We incremented p_{advanced} one percentage point at a time, and recorded the values of the seven query and result-click features at each point. The values of the features at four milestones ($> 0\%$, $\geq 25\%$, $\geq 50\%$, and $\geq 75\%$) are shown in Table 4. As can be seen in the table, as p_{advanced} increases, differences in the features between those using advanced syntax and those not using advanced syntax become more substantial. However, it is interesting to note that as p_{advanced} increases, the number of queries submitted per day actually falls (Pearson's $R = -.512$, $t(98) = 5.98$, $p < .0001$). More advanced users may need to pose fewer queries to find relevant information.

To study the patterns of relationship among these dependent variables (including the p_{advanced}), we applied *factor analysis* [26].

Table 5 shows the intercorrelation matrix between the features and the percentage of queries with operators (P_{advanced}). Each cell in the table contains the Pearson’s correlation coefficient between the two features for a given row-column pair.

Table 5. Intercorrelation matrix (query / result-click features).

| | p_{adv} | QPS | QRR | QWL | QPD | ACP | CP | ASC |
|------------------|------------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| p_{adv} | 1.00 | .946 | .970 | .987 | -.512 | .930 | -.746 | -.583 |
| QPS | | 1.00 | .944 | .943 | -.643 | .860 | -.594 | -.712 |
| QRR | | | 1.00 | .934 | -.462 | .919 | -.621 | -.667 |
| QWL | | | | 1.00 | -.392 | .612 | -.445 | .735 |
| QPD | | | | | 1.00 | .676 | .780 | .943 |
| ACP | | | | | | 1.00 | .838 | .711 |
| CP | | | | | | | 1.00 | .654 |
| ASC | | | | | | | | 1.00 |

It is only the first data column and row that reflect the correlations between p_{advanced} and the other query and result-click features. Columns 2 – 8 show the inter-correlations between the other features. There are strong positive correlations between some of the features (e.g., the number of words in the query (QWL) and the average probability of clicking on a search result (ACP)). However, there were also fairly strong negative correlations between some features (e.g., the average length of the queries (QWL) and the probability of clicking on a search result (CP)). The factor analysis revealed the presence of two factors that account for 83.6% of the variance. As is standard practice in factor analysis, all features with an absolute factor loading of .30 or less were removed. The two factors that emerged, with their respective loadings, can be expressed as:

$$\begin{aligned} \text{Factor A} &= .98(\text{QRR}) + .97(p_{\text{adv}}) + .97(\text{QPS}) \\ &\quad + .71(\text{ACP}) + .69(\text{QWL}) \\ \text{Factor B} &= .96(\text{CP}) + .90(\text{QPD}) + .67(\text{ACP}) + .52(\text{ASC}) \end{aligned}$$

Variance in the query and result-click behavior of our advanced search engine users can be expressed using these two constructs. Factor A is the most powerful, contributing 50.5% of the variance. It appears to represent a very basic dimension of variance that covers *query attributes and querying behavior*, and suggests a relationship between query properties (length, frequency, complexity, and repetition) and the position of users’ clicks in the result list. The dimension underlying Factor B accounts for 33.1% of the variance, and describes attributes of *result-click behavior*, and a strong correlation between result clicks and the number of queries submitted each day.

Summary: In this section we have shown that there are marked differences in aspects of the querying and result-clickthrough behaviors of advanced users relative to non-advanced users. We have also shown that the greater the proportion of queries that contain advanced syntax, the larger the differences in query and clickthrough behaviors become. A factor analysis revealed the presence of two dimensions that adequately characterize variance in the query and result-click features. In the *querying* dimension query attributes, such as the length and proportion that contain advanced syntax, and querying behavior, such as the number of queries submitted per day both affect result-click position. In addition, in the *result-click* dimension, it appears that daily querying frequency influences result-click features such as the

likelihood that a user will click on a search result and the amount of time between result presentation and the search result click.

The features used in this section are only interactions with search engines in the form of queries and result clicks. We did not address how users searched for information beyond the result page. In the next section we use the search trails described in Section 4 to analyze the post-query browsing behavior of users.

5.2 Post-query browsing behavior

In this section we look at several attributes of the search trails users followed beyond the results page in an attempt to discern whether the use of advanced search syntax can be used as a predictor of aspects of post-query interaction behavior.

As we did previously, we first describe the mean average values for each of the browsing features, across all advanced users (i.e. $P_{\text{advanced}} > 0\%$), all non-advanced users (i.e., $P_{\text{advanced}} = 0\%$), and all users regardless of their estimated search expertise level. We then look at the effect on the browsing features of increasing the value of P_{advanced} required to be considered “advanced” from 1% to 100%. In Table 6 we present the average values for each of these features for the two groups of users. Also shown are the percentage of search trails ($\%_{\text{Trails}}$) and the percentage of users ($\%_{\text{Users}}$) used to compute the averages.

Table 6. Post-query browsing features (per trail).

| Feature | P_{advanced} | | | | |
|----------------------|-----------------------|---------------|-------------|-------------|-------------|
| | 0% | > 0% | $\geq 25\%$ | $\geq 50\%$ | $\geq 75\%$ |
| Session secs. | 701.10 | 706.21 | 792.65 | 903.01 | 1114.71 |
| Trail secs. | 205.39 | 159.56 | 156.45 | 147.91 | 136.79 |
| Display secs. | 36.95 | 32.94 | 34.91 | 33.11 | 30.67 |
| Num. steps | 4.88 | 4.72 | 4.40 | 4.40 | 4.39 |
| Num. backs | 1.20 | 1.02 | 1.03 | 1.03 | 1.02 |
| Num. branches | 1.55 | 1.51 | 1.50 | 1.47 | 1.44 |
| $\%_{\text{Trails}}$ | 72.14% | 27.86% | .83% | .23% | .05% |
| $\%_{\text{Users}}$ | 79.90% | 20.10% | .79% | .18% | .04% |

As can be seen from Table 6, there are differences in the post-query interaction behaviors of advanced users ($p_{\text{advanced}} > 0\%$) relative to that do not use query operators in any of their queries ($p_{\text{advanced}} = 0\%$). Once again, the columns of interest in this comparison are bolded. As we did in Section 5.1 for query and result-click behavior, we performed an independent measures t -test between the values reported for each of the post-query browsing features. The results of this test suggest that differences between those that use advanced syntax and those that do not are significant ($t(12495029) \geq 3.09$, $p \leq .002$, $\alpha = .008$). Given the sample sizes, all of the differences between means in the two groups were significant. However, we once again applied a Cohen’s d -test to determine the effect size. The findings (ranked in descending order based on effect size), show that relative to non-advanced users, advanced search engine users:

- Revisit pages in the trail less often ($d = .45$)
- Spend less time traversing each search trail ($d = .38$)
- Spend less time viewing each document ($d = .28$)
- “Branch” (i.e., proceed to new pages following a back operation) less often ($d = .18$)
- Follow search trails with fewer steps ($d = .16$)

It seems that advanced users use a more directed searching style than non-advanced users. They spend less time following search trails and view the documents that lie on those trails for less time. This is in accordance with our earlier proposition that advanced users seem able to discern document relevance in less time. Advanced users also tend to deviate less from a direct path as they search, with fewer revisits to previously-visited pages and less branching during their searching.

As we did in the previous section, we increased the p_{advanced} threshold one point at a time. With the exception of number of back operations (NB), the values attributable to each of the features change as p_{advanced} increased. It seems that the differences noted earlier between non-advanced users and those that use any advanced syntax become more significant as p_{advanced} increases. As in the previous section, we conducted a factor analysis of these features and p_{advanced} . Table 7 shows the intercorrelation matrix for all these variables.

Table 7. Intercorrelation matrix (post-query browsing).

| | p_{adv} | SS | TS | DS | NS | NB | NBA |
|------------------|------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| p_{adv} | 1.00 | .977 | -.843 | -.867 | -.395 | -.339 | -.249 |
| SS | | 1.00 | -.765 | -.875 | -.374 | -.335 | -.237 |
| TS | | | 1.00 | .948 | .387 | .281 | .250 |
| DS | | | | 1.00 | .392 | .344 | .257 |
| NS | | | | | 1.00 | .891 | .934 |
| NB | | | | | | 1.00 | .918 |
| NBA | | | | | | | 1.00 |

As the proportion of queries containing advanced syntax increases, the values of many of the post-query browsing features decrease. Only the average session time (SS) exhibits a strong positive correlation with p_{advanced} . The factor analysis revealed the presence of two factors that account for 89.8% of the variance. Once again, all features with an absolute factor loading of .30 or less were removed. The two factors that emerged, with their respective loadings, can be expressed as:

$$\text{Factor A} = .95(\text{DS}) + .88(\text{TS}) - .91(\text{SS}) - .95(p_{\text{adv}})$$

$$\text{Factor B} = .99(\text{NBA}) + .93(\text{NS}) + .91(\text{NB})$$

Variance in the query and result-click behavior of those who use query operators can be expressed using these two constructs. Factor A is the most powerful, contributing 50.1% of the variance. It appears to represent a very basic *temporal dimension* that covers timing and percentage of queries with advanced syntax, and suggests a negative relationship between time spent searching and overall session time, and a negative relationship between time spent searching and p_{advanced} . The *navigation dimension* underlying Factor B accounts for 39.7% of the variance, and describes attributes of post-query navigation, all of which seem to be strongly correlated with each other but not p_{advanced} or timing.

Summary: In this section we have shown that advanced users' post-query browsing behavior appears more directed than that of non-advanced users. Although their search sessions are longer, advanced users follow fewer search trails during their sessions, (i.e., submit fewer queries), their search trails are shorter, and their trails exhibit fewer deviations or regressions to previously encountered pages. We also showed that as p_{advanced} increases, session time increases (perhaps more advanced users are

multitasking between search and other operations), and search interaction becomes more focused, perhaps because advanced users are able target relevant information more effectively, with less need for regressions or deviations in their search trails.

As well as interaction behaviors such as queries, result clicks, and post-query browse behavior, another important aspect of the search process is the attainment of information relevant to the query. In the next section we analyze the success of advanced and non-advanced users in obtaining relevant information.

5.3 Search success

As described earlier, we used six-level relevance judgments assigned to query-document pairs as an approximate measure of search success based on documents encountered on search trails. However, the queries for which we have judgments generally did not contain advanced operators. To maximize the likelihood of coverage we removed advanced operators from all queries when retrieving the relevance judgments. The mean average relevance judgment values for each of the four metrics – first, last, average, and maximum – are shown in Table 8 for non-advanced users (0%) and advanced users (>0%).

Table 8. Search success (min. = 1, max. = 6) (per trail).

| Feature | | p_{advanced} | | | | |
|---------|-----------|-----------------------|------|-------|-------|-------|
| | | 0% | > 0% | ≥ 25% | ≥ 50% | ≥ 75% |
| First | <u>M</u> | 4.03 | 4.19 | 4.24 | 4.26 | 4.57 |
| | <u>SD</u> | 1.58 | 1.56 | 1.34 | 1.38 | 1.27 |
| Last | <u>M</u> | 3.79 | 3.92 | 4.00 | 4.13 | 4.35 |
| | <u>SD</u> | 1.60 | 1.57 | 1.29 | 1.25 | .89 |
| Max. | <u>M</u> | 4.04 | 4.20 | 4.19 | 4.19 | 4.46 |
| | <u>SD</u> | 1.63 | 1.51 | 1.28 | 1.37 | 1.25 |
| Avg. | <u>M</u> | 3.93 | 4.06 | 4.08 | 4.08 | 4.26 |
| | <u>SD</u> | 1.57 | 1.51 | 1.23 | 1.32 | 1.14 |

The findings suggest that users who use advanced syntax at all ($p_{\text{advanced}} > 0\%$) were more successful – across all four measures – than those who never used advanced syntax ($p_{\text{advanced}} = 0\%$). Not only were these users more successful in their searching, but they were *consistently* more successful (i.e., the standard deviation in relevance scores is lower for advanced users and continues to drop as p_{advanced} increases). The differences in the four mean average relevance scores for each metric between these two user groups were significant with independent measures *t*-tests (all $t(516765) \geq 3.29$, $p \leq .001$, $\alpha = .0125$). As we increase the value of p_{advanced} as in previous sections, the average relevance score across all metrics also increases (all Pearson's $R \geq .654$), suggesting that *more* advanced users are also *more* likely to succeed in their searching. The searchers that use advanced operators may have additional skills in locating relevant information, or may know where this information resides based on previous experience.³ Despite the fact that the four metrics targeted different parts of the search trail (e.g., first vs. last) or different ways to gather relevant information (e.g., average vs. maximum), the differences between groups and within the advanced group were consistent.

³ Although in our logs there was no obvious indication of more revisitation by advanced search engine users.

To see whether there were any differences in the nature of the queries submitted by advanced search engine users, we studied the distribution of the four advanced operators: quotation marks, plus, minus, and “site:”. In Table 9 we show how these operators were distributed in all queries submitted by these users.

Table 9. Distribution of query operators.

| Feature | P _{advanced} | | | |
|---------------------|-----------------------|-------|-------|-------|
| | > 0% | ≥ 25% | ≥ 50% | ≥ 75% |
| Quotes (“”) | 71.08 | 77.09 | 70.33 | 70.00 |
| Plus (+) | 6.84 | 13.31 | 19.21 | 33.90 |
| Minus (-) | 6.62 | 2.88 | 1.96 | 2.42 |
| Site: | 21.55 | 12.72 | 13.04 | 9.86 |
| Avg. num. operators | 1.08 | 1.14 | 1.28 | 1.49 |

The distribution of the quotes, plus, and minus operators are similar amongst the four levels of P_{advanced}, with quotes being the most popular of the four operators used. However, it appears that the plus operator is the main differentiator between the P_{advanced} user groups. This operator, which forces the search engine to include in the query terms that are usually excluded by default (e.g. “the”, “a”), may account for some portion of the difference in observed search success.⁴ However, this does not capture the contribution that each of these operators makes to the increase in relevance compared with excluding the operator. To gain some insight into this, we examined the impact that each of the operators had on the relevance of retrieved results. We focused on queries in P_{advanced} > 0% where the same user had issued a query without operators and the same query with operators either before or afterwards. Although there were few queries with matching pairs – and almost all of them contained quotes – there was a small (approximately 10%) increase in the average relevance judgment score assigned to documents on the trail with quotes in the initial query. It may be the case that quoted queries led to retrieval of more relevant documents, or that they better match the perceived needs of relevance judges and therefore lead to judged documents receiving higher scores. More analysis similar to [8] is required to test these propositions further.

Summary: In this section we have used several measures to study the search success of advanced and non-advanced users. The findings of our analysis suggest that advanced search engine users are more successful and have more consistency in the relevance of the pages they visit. Their additional search expertise may make these users better able to make better decisions about which documents to view, meaning they encounter consistently more relevant information on their searches. In addition, within the group of advanced users there is a strong correlation between P_{advanced} and the degree of search success. Advanced search engine users may be more adept at combining query operators to formulate powerful query statements. We now discuss the findings from all three subsections and their implications for the design of improved Web search systems.

⁴ It is worth noting that there were no significant differences in the distribution of usage of the three search engines – Google, Yahoo!, or Windows Live Search – amongst advanced search engine users, or between advanced users and non-advanced.

6. DISCUSSION AND IMPLICATIONS

Our findings indicate significant differences in the querying, result-click, post-query navigation, and search success of those that use advanced syntax versus those that do not. Many of these findings mirror those already found in previous studies with groups of self-identified novices and experts [13][19]. There are several ways in which a commercial search engine system might benefit from a quantitative indication of searcher expertise. This might be yet another feature available to a ranking engine; i.e. it may be the case that expert searchers in some cases prefer different pages than novice searchers. The user interface to a search engine might be tailored to a user’s expertise level; perhaps even more advanced features such as term weighting and query expansion suggestions could be presented to more experienced searchers while preserving the simplicity of the basic interface for novices. Result presentation might also be customized based on search skill level; future work might re-evaluate the benefits of content snippets, thumbnails, etc. in a manner that allows different outcomes for different expertise levels. Additionally, if browsing histories are available, the destinations of advanced searchers could be used as suggested results for queries, bypassing and potentially improving upon the traditional search process [10].

The use of the interaction of advanced search engine users to guide others with less expertise is an attractive proposition for the designers of search systems. In part, these searchers may have more post-query browsing expertise that allows them to overcome the shortcomings of search systems [29]. Their interactions can be used to point users to places that advanced search engine users visit [32] or simply to train less experienced searchers how to search more effectively. However, if expert users are going to be used in this way, issues of data sparsity will need to be overcome. Our advanced users only accounted for 20.1% of the users whose interactions we studied. Whilst these may be amongst the most active users it is unlikely that they will view documents that cover large number of subject areas. However, rather than focusing on where they go (which is perhaps more appropriate for those with domain knowledge), advanced search engine users may use moves, tactics and strategies [2] that inexperienced users can learn from. Encouraging users to use advanced syntax helps them learn how to formulate better search queries; leveraging the searching style of expert searchers could help them learn more successful post-query interactions.

One potential limitation to the results we report is that in prior research, it has been shown that query operators do not significantly improve the effectiveness of Web search results [8], and that searchers may be able to perform just as well without them [27]. It could therefore be argued that the users who *do not* use query operators are in fact more advanced, since they do not waste time using potentially redundant syntax in their query statements. However, this seems unlikely given that those who use advanced syntax exhibited search behaviors typical of users with expertise [13], and are more successful in their searching. However, in future work we will expand of definition of “advanced user” beyond attributes of the query to also include other interaction behaviors, some of which we have defined in this study, and other avenues of research such as eye-tracking [12].

7. CONCLUSIONS

In this paper we have described a log-based study of search behavior on the Web that has demonstrated that the use of advanced search syntax is correlated with other aspects of search

behavior such as querying, result clickthrough, post-query navigation, and search success. Those that use this syntax are active online for longer, spend less time querying and traversing search trails, exhibit less deviation in their trails, are more likely to explore search results, take less time to click on results, and are more successful in their searching. These are all traits that we would expect expert searchers to exhibit. Crude classification of users based on just one feature that is easily extractable from the query stream yields remarkable results about the interaction behavior of users that do not use the syntax and those that do. As we have suggested, search systems may leverage the interactions of these users for improved document ranking, page recommendation, or even user training. Future work will include the development of search interfaces and modified retrieval engines that make use of these information-rich features, and further investigation into the use of these features as indicators of search expertise, including a cross-correlation analysis between result click and post-query behavior.

8. ACKNOWLEDGEMENTS

The authors are grateful to Susan Dumais for her thoughtful and constructive comments on a draft of this paper.

9. REFERENCES

- [1] Anick, P. (2003). Using terminological feedback for Web search refinement: A log-based study. In *Proc. ACM SIGIR*, pp. 88-95.
- [2] Bates, M. (1990). Where should the person stop and the information search interface start? *Inf. Proc. Manage.* 26, 5, 575-591.
- [3] Belkin, N.J. (2000). Helping people find what they don't know. *Comm. ACM*, 43, 8, 58-61.
- [4] Belkin, N.J. et al. (2003). Query length in interactive information retrieval. In *Proc. ACM SIGIR*, pp. 205-212.
- [5] Bhavnani, S.K. (2001). Domain-specific search strategies for the effective retrieval of healthcare and shopping information. In *Proc. ACM SIGCHI*, pp. 610-611.
- [6] Chi, E. H., Pirolli, P. L., Chen, K. & Pitkow, J. E. (2001). Using information scent to model user information needs and actions and the Web. In *Proc. ACM SIGCHI*, pp. 490-497.
- [7] De Lima, E.F. & Pedersen, J.O. (1999). Phrase recognition and expansion for short, precision-biased queries based on a query log. In *Proc. of ACM SIGIR*, pp. 145-152.
- [8] Eastman, C.M. & Jansen, B.J. (2003). Coverage, relevance, and ranking: The impact of query operators on Web search engine results. *ACM TOIS*, 21, 4, 383-411.
- [9] Efthimiadis, E.N. (1996). Query expansion. *Annual Review of Information Science and Technology*, 31, 121-187.
- [10] Furnas, G. (1985). Experience with an adaptive indexing scheme. In *Proc. ACM SIGCHI*, pp. 131-135.
- [11] Furnas, G.W., Landauer, T.K., Gomez, L.M. & Dumais, S.T. (1987). The vocabulary problem in human-system communication: An analysis and a solution. *Comm. ACM*, 30, 11, 964-971.
- [12] Granka, L., Joachims, T. & Gay, G. (2004). Eye-tracking analysis of user behavior in WWW search. In *Proc. ACM SIGIR*, pp. 478-479.
- [13] Hölscher, C. & Strube, G. (2000). Web search behavior of internet experts and newbies. In *Proc. WWW*, pp. 337-346.
- [14] Jansen, B.J. (2000). An investigation into the use of simple queries on Web IR systems. *Inf. Res.* 6, 1.
- [15] Jansen, B.J., Spink, A. & Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the Web. *Inf. Proc. Manage.* 36, 2, 207-227.
- [16] Jones, R., Rey, B., Madani, O. & Greiner, W. (2006). Generating query substitutions. In *Proc. WWW*, pp. 387-396.
- [17] Kaski, S., Myllymäki, P. & Kojo, I. (2005). User models from implicit feedback for proactive information retrieval. In *Workshop at UM Conference; Machine Learning for User Modeling: Challenges*.
- [18] Kelly, D., Dollu, V.D. & Fu, X. (2005). The loquacious user: a document-independent source of terms for query expansion. In *Proc. ACM SIGIR*, pp. 457-464.
- [19] Lazonder, A.W., Biemans, H.J.A. & Woperis, I.G.J.H. (2000). Differences between novice and experienced users in searching for information on the World Wide Web. *J. ASIST*. 51, 6, 576-581.
- [20] Morita, M. & Shinoda, Y. (1994). Information filtering based on user behavior analysis and best match text retrieval. In *Proc. ACM SIGIR*, pp. 272-281.
- [21] NIST Special Publication 500-266: The Fourteenth Text Retrieval Conference Proceedings (TREC 2005).
- [22] Oddy, R. (1977). Information retrieval through man-machine dialogue. *J. Doc.* 33, 1, 1-14.
- [23] Rose, D.E. & Levinson, D. (2004). Understanding user goals in Web search. In *Proc. WWW*, pp. 13-19.
- [24] Salton, G. and Buckley, C. (1990). Improving retrieval performance by relevance feedback. *J. ASIST*, 41 4, 288-287.
- [25] Silverstein, C., Marais, H., Henzinger, M. & Moricz, M. (1999). Analysis of a very large web search engine query log. *SIGIR Forum*, 33, 1, 6-12.
- [26] Spearman, C. (1904). "General intelligence," objectively determined and measured. *Amer. J. Psy.* 15, 201-293.
- [27] Spink, A., Bateman, J. & Jansen, B.J. (1998). Searching heterogeneous collections on the Web: Behavior of Excite users. *Inf. Res.* 4, 2, 317-328.
- [28] Spink, A., Griesdorf, H. & Bateman, J. (1998). From highly relevant to not relevant: examining different regions of relevance. *Inf. Proc. Manage.* 34 5, 599-621.
- [29] Teevan, J. et al. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proc. ACM SIGCHI*, pp. 415-422.
- [30] Teevan, J., Dumais, S.T. & Horvitz, E. (2005). Personalizing search via automated analysis of interests and activities. In *Proc. ACM SIGIR*, pp. 449-456
- [31] Wang, P., Berry, M. & Yang, Y. (2003). Mining longitudinal Web queries: Trends and patterns. *J. ASIST*, 54, 3, 742-758.
- [32] White, R.W., Bilenko, M. & Cucerzan, S. (2007). Studying the use of popular destinations to enhance Web search interaction. In *Proc. ACM SIGIR*, in press.
- [33] White, R.W. & Drucker, S. (2007). Investigating behavioral variability in Web search. In *Proc. WWW*, in press.
- [34] White, R.W., Ruthven, I. & Jose, J.M. (2002). Finding relevant documents using top-ranking sentences: An evaluation of two alternative schemes. In *Proc. ACM SIGIR*, pp. 57-64.
- [35] Wildemuth, B.M., do Bleik, R., Friedman, C.P. & File, D.D. (1995). Medical students' personal knowledge. Search proficiency, and database use in problem solving. *J. ASIST*, 46, 590-607.