How does Search Behavior Change as Search Becomes More Difficult?

Anne Aula, Rehan M. Khan and Zhiwei Guan
Google
1600 Amphitheatre Parkway, Mountain View, CA 94043
{anneaula, rkhan, zguan}@google.com

ABSTRACT
Search engines make it easy to check facts online, but finding some specific kinds of information sometimes proves to be difficult. We studied the behavioral signals that suggest that a user is having trouble in a search task. First, we ran a lab study with 23 users to gain a preliminary understanding on how users' behavior changes when they struggle finding the information they're looking for. The observations were then tested with 179 participants who all completed an average of 22.3 tasks from a pool of 100 tasks. The large-scale study provided quantitative support for our qualitative observations from the lab study. When having difficulty in finding information, users start to formulate more diverse queries, they use advanced operators more, and they spend a longer time on the search result page as compared to the successful tasks. The results complement the existing body of research focusing on successful search strategies.

Author Keywords
Web search, search engines, difficult search tasks, search strategies, behavioral signals.

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors.

INTRODUCTION
Studies of search behavior have often focused on the differences between the search strategies of novices and experts. By studying experts, researchers have hoped to understand successful strategies in information search and conversely, observing the behavior of novice searchers, strategies that are less successful. However, as we will

explain later in this paper, the definitions for experts and novices have varied widely between different studies and it is somewhat unclear how strategies revealed in these studies actually relate to success in search tasks.

Our experience observing hundreds of users in field and usability lab studies suggest that even highly skillful searchers sometimes struggle trying to find the information they're looking for. When that happens, searchers often try a number of different queries, they get frustrated, and finally, they give up and decide to find the information some other way (e.g., asking a friend). Before they give up, there are observable changes in their behavior. Their body language changes (e.g., they frown, some start biting their nails, and many lean closer to the monitor as if to make sure they're not missing something obvious), they start sighing, and in think-aloud studies, they tend to forget to think-aloud. These changes are easy for a human observer to recognize. However, these behaviors - or the underlying frustration - seem to be linked to behavioral changes that could potentially be detected by the computer, too.

A number of studies have indirectly compared successful and less successful search strategies by comparing expert and novice searchers in lab studies. Recently, researchers have also begun to use data from search engine logs to identify metrics that are related to users' search success. These studies have provided some promising findings, but the noisiness of the log data makes it hard to determine if the searchers were successful or not and which signals are specific to which kinds of tasks.

Rather than studying successful strategies, our focus is on failures. What happens when the searcher is facing serious difficulties finding a piece of information? What are the signals we could use to identify user frustration? Our approach combines small-scale lab studies and log analyses: first we gained a qualitative understanding of how users' search behavior changes when they start having difficulties in search tasks and generated hypotheses on these changes could be quantified. Then we tested the hypotheses with a large-scale online study. Importantly, we focus on closed informational search tasks where search success is easy to measure. By analyzing users' behavior in tasks they fail on and comparing these behaviors to their

behavior in tasks where they succeed, we hope to identify signals in user behavior that suggest that the searcher is likely to be frustrated. This information will be extremely valuable to search engine providers, who could recognize frustration in real time and potentially offer suggestions.

RELATED RESEARCH

Based on search logs, the average number of query terms is between 2.35 terms and 2.6 terms per query [13, 21, 23] - a more recent study reported a slightly higher number, 2.93 terms per query (for queries issued from computers or iPhones) [16]. Most of these queries are simple keyword queries – only about 10% of queries contain advanced query operators [14, 23]. However, there are large regional differences in the use of advanced operators (they are used more in the U.S. than in Europe) [12]. An analysis by Eastman and Jansen suggested that most of the query operators do not increase the precision of the query so they might not be worth the trouble [7]. After typing in the query, search engine users typically evaluate the results quickly before they either click on a result or refine their query (an average of 7.78 seconds reported in [9]); the time used to evaluate the results seems to increase with the difficulty of the task. Typical web search sessions are short, containing less than two queries, on average [16].

In addition to the average numbers describing web search behavior, researchers have focused on the interaction flow. White and Drucker [25] studied behavioral variability in web search. They focused on users' queries, time and structure of interactions, and the types of webpages users visited. Their analysis showed two extreme user groups: explorers and navigators. Navigators' interaction is highly consistent in the trails they follow (e.g., tackle problems serially, more likely to revisit domains) whereas explorers' search interaction is highly variable (e.g., many queries, visiting many domains). Most of their users (80%) were somewhere in between showing both variability and consistency in their search interaction. The authors expect the search styles to vary by task: they assume searchers will exhibit navigator type systematic behaviors in well-defined fact-finding tasks and explorer-type more variable behavior in complex sense-making tasks.

A large body of research has focused on the effect of expertise on search strategies. Most often, the studies have been small-scale lab studies. These assume that the "expert strategies" are those that lead to higher success in search tasks. In these studies experts, as compared to novices, tend to spend more [5] or less [22] time in search tasks, reformulate the queries less often [11], use query formatting tools more often and make less errors using them [4, 11], use longer queries [1, 11], use a "depth-first" or "bottom-up" strategy [15, 19], use a more systematic query refinement strategy [8], and have a similar [5] or higher level of performance [17, 18].

The studies focusing on the differences between experts and novices have used widely different definitions of what

makes one an expert: more than 5 years of computer and more than 4.5 years of web use [15]; more than 5 hours of browsing a week [17]; over 50 hours of www experience and high self-reported proficiency [18]; and at least of 3 years of extensive professional experience [11] - all these criteria have been used to define experts. In addition to the definition of an expert, the metrics used to measure success have also varied significantly between the studies. Thus, the relationship between these strategies and how they relate to the success in search tasks is unclear.

Instead of grouping users into experts and novices, Aula and Nordhausen [3] focused on different behavioral strategies that explained participants' success in web search tasks. Their study suggested that in fact-finding tasks, the speed of querying was related to an increase in search success (as measured by task-completion speed - the number of tasks successfully completed per time unit). In essence, a high speed of querying means that the user formulates queries and evaluates the search results quickly. Their study also suggested that users often started well-defined fact-finding tasks with general queries using only selected keywords from the task description and only after a couple of queries, ended up using more precise - natural-language like queries which often ended up being successful. In their study, more successful searchers seemed to be more systematic in their query formulation which often meant that they only changed their query slightly from refinement to refinement whereas less successful searchers' refinement strategy was more random-looking.

White and Morris [26] analyzed search logs of different search engines to shed light on how advanced searchers differ from their less advanced counterparts. They defined "advanced search engine users" as users who use advanced operators in their queries. When studying the differences in the search logs of those users who used operators and those who didn't, they found several differences in their behavior. For example, advanced users query less frequently in a session, they compose longer queries, and they click further down the result list as compared to non-advanced users. After formulating the query, advanced users are more directed than non-advanced users - they submit fewer queries in their sessions, their search trails are shorter and they deviate from the search trails less often than non-advanced users. The authors also measured the differences in the success rate of advanced and non-advanced users by having external judges rate the query-webpage pairs from the users' logs. Their analysis showed that advanced users were generally more successful than non-advanced users in that they visited more relevant pages.

Huntington et al. [10] analyzed BBC search engine log files and found some evidence that time lapse between searches and the number of searches in a session could be used to evaluate the effectiveness of search engine functionality and consequently, search success. Downey et al. [6] analyzed more than 80 million URL visits. They found that for rare search goals, submitting a common query to the

search engine and using hyperlinks to navigate to a page that contains the information is a more effective approach than submitting a rare query. This is similar to the "orienteering" approach described by [24].

Earlier research has mainly focused on the differences between experts and novices or mined the log data to find metrics that correlate with success. Instead of looking at individual differences (by pre-categorizing users into experts or novices) or having to rely only on log data where the level of success is difficult to confirm, we used the users' own perception of success or failure to explain the changes in their behavior.

METHODS

Experiment #1: Usability lab study

Participants
23 participants were recruited via an online ad and from our participant database. All the participants used Google as their primary search engine and all of them used search engines at least several times a week. The participants received gift checks as compensation for their time.

Procedure
The participants did the study in our usability lab using a PC running Windows XP and with a screen resolution of 1024x768. The browser they used was Internet Explorer 8. We gave each participant two or three difficult search tasks and a varying number of easier filler tasks (depending on how much time was left) in a pseudo-random order. We always started the session with an easy task and in between two difficult tasks, there was always at least one easy task. The last task was always easy to make the session end on a positive note.

All tasks (both the easy filler tasks and the difficult tasks) were directed closed informational tasks ("I want to get an answer to a question that has a single, unambiguous answer" [20]). We chose the difficult tasks so that they seemed as simple as the easy tasks. We did not want the participants to approach these tasks differently just because they appeared difficult – instead, we wanted to see what happened when they realized that the task was difficult while they were working on it. The two difficult tasks and a couple of easy tasks we used were:

- You once heard that the Dave Matthews Band owns a studio in Virginia but you don't know the name of it. The studio is located outside of Charlottesville and it's in the mountains. What is the name of the studio? (difficult)

- I was recently watching footage of one of Prada's fashion shows from a few months ago, where two models fell (and several others stumbled) due to the footwear. Find the names of the models who fell. (difficult)

- I was watching the movie "Stand by Me" the other day. I know it is based on a Stephen King story with a different name. What is the name of the story? (easy)

- My friend had a cool program on his iphone that told him what song was playing when he held the phone to the speaker playing the song. I forgot to ask him but really want to get a copy for my own iPhone. What is the name of this program? (easy)

The participants were asked to start the tasks with Google, but they were told that after that, they could use whatever websites or search engines they would typically use if they were searching for this information at home. Overall, they were encouraged to search as they normally would if they were at home searching for this information for their own use. We did not use a think-aloud protocol to avoid distracting or slowing down the searchers and to keep the setting as natural as possible.

After finding the answer to the task, the users were instructed to highlight the answer on the webpage or in the search results with their mouse and press a specific key on the keyboard to close the browser window and stop the recording. We didn't require the participant to find the correct answer to the task and we didn't tell them whether their answer was correct - they were simply asked to find an answer they believed to be the correct. If they couldn't find the answer, the participants were instructed to tell the moderator that they wanted to move on to the next task and then press the key on the keyboard to stop the recording.

Data collection
We logged all the URLs the users visited during the session with timestamps. We also recorded what was happening on the screen and the users' voice during the tasks. After the sessions, we added the task success rating (successful/gave up) to the log files so that we could analyze these tasks separately.

Experiment #2: Online study

Participants
179 participants (18-54 years old) took part in this study. The participants were given monetary compensation for their time.

Procedure
Each participant was randomly given an average of 22.3 tasks (interquartile range is 8 and 31) from the pool of 100 search tasks of varying difficulty, with the constraint that no task was completed by more than 40 people (the participants could do as many tasks as they liked). The 100 tasks we used for the study were closed informational tasks (the same task type we used in the lab study). The participants used their own computer for the study. Users were required to use FireFox as the web browser.

When starting the task, the participants were first shown the task description. Examples of tasks are listed below:

- Find a map drawing of the flight path of a space shuttle that flew in 2008.

- Who is the athletic director at Mater Dei High School in Santa Ana California?

After reading the task description, the participants clicked a start button and were taken to a search engine page to start the search task. Participants were told they should continue the task until completed or until 7 minutes had elapsed. After finding the answer (or abandoning the task), the participants were taken to a response page in which they could enter their answer, and indicated whether they had succeeded with the task and rated their satisfaction with their experience on a 5-point Likert scale. The time to complete each task was recorded.

Data collection
We logged the URLs users visited during the session along with timestamps. We also had each user's answer to each task they completed, whether or not they thought they had succeeded, and their rating of the task.

RESULTS
We report the usability lab study findings first followed by the online study findings. The analysis of the laboratory study focuses on qualitative findings that were used as hypothesis to be tested with the large-scale online study.

Experiment #1: Usability lab study

Typically, the queries participants formulated for the closed informational tasks contained the main facets from the task description. In the easier filler tasks, this approach was mostly successful. However, in the tasks where they struggled to find information, participants often resorted to a different strategy after a number of failed attempts with keywords: they asked a direct question. Below, we show how two users changed their strategy from keywords to direct questions. User A:

- [prada fashion show models falling]

- [prada fashion show models name falling]

- [prada fashion model name]

- [prada fashion model name fall]

- [prada fashion model name falling]

- [prada fashion models names that fell]

- [prada models fall]

- [what were the names of the models that fell at prada]

User B:
- [Dave matthews band studio Virginia outside charlottesville]

- [Dave matthews band name studio Virginia outside charlottesville]

- [What is the name of studio in Virginia "Dave matthews band"?]

- [What is the name of studio in charlottesville "Dave matthews band"?]

Unfortunately, most of the time, the question queries failed to give users the information they were looking for. After trying the question approach, some users went back to trying keyword queries and some gave up.

Another interesting finding related to query formulation is the way users refine their queries. Earlier research has suggested that less successful searchers tend to be less systematic in their query refinement process [3]. Our study suggests that this unsystematic refinement process might more generally indicate that the user is having difficulty with the search task (rather than necessarily being a "strategy" the user employs in all the search tasks). Many users picked an initial approach and only made subtle changes to the query with each refinement. Sometimes they added, removed, or changed a term in the query - or they tried using advanced operators, such as quotes. However, if the small changes to the query weren't enough, they often ended up changing how they approached the task. It seemed that with increased frustration, the users ended up changing their approach several times.

Below is an example of the queries one participant formulated for a task where the goal was to find out which US president named his child after the child's grandfather's college buddy. This participant picked an approach (trying to find a list of presidents), made subtle changes to the queries, realized this approach would not work, changed the approach (find more information about George Bush), made subtle changes, changed the approach again (started asking questions and using phrase search), etc.

- [US presidents late 1900]

- [US presidents]

- [time line US presidents]

- [timeline US presidents]

- [george bush wiki]

- [george bush children names]

- [george H. bush children names]

- [who did george H. bush name his child after?]

- [president's child name after college buddy]

- [president's child name after "college buddy"]

- [US president's child name after "college buddy"]

- [US president child named after "college buddy"]

- [college buddy "US president"]

- [college buddy "US president" child]

- [which US President name child after college buddy?]

- [which US President late 1900s name child after college buddy?]

- [US President late 1900s]

- [US President with children]

In successful tasks the query refinement process seemed to be much more straightforward. Oftentimes, the users started with a more general query and made the query more specific (longer) with each refinement until eventually, they found the information.

- [chelsea hotel singer]

- [chelsea hotel singer name]

- [leonard cohen chelsea hotel song background]

Users spend, on average, about 8 seconds on the results page before selecting a result or refining their query, and for hard tasks they spend slightly longer (11 seconds in [9]). In our study, in the tasks where users gave up, the time they spent on the results page was often significantly longer than the typical time reported by Granka et al. [9] - there were even cases where the user ended up spending over a minute on the search result page. During that time, users scanned the results and other options on the page, and sometimes they started to refine the query (they clicked on the search box and even typed in something), but they could not think of a better query so they never submitted the query.

Based on this data, we formulated the following hypotheses: when users are having difficulties in a search task, they will

- spend more time on the search result page,

- use more natural language/question type queries and/or use advanced operators in their queries, and

- have a more unsystematic query refinement process.

	All tasks	Successful tasks	Unsuccessful tasks
Average time on task	223.9 (2.36)	176.2 2.24	384.6 (3.52)
Average number of query terms/query	4.77 (0.029)	4.66 (0.030)	5.13 (0.027)
Average number of queries/task	6.71 (0.098)	4.98 (0.070)	12.41 (0.098)
Proportion of queries with advanced operators ('+', '-', 'AND', 'OR', ':')	0.074 (0.0024)	0.056 (0.0046)	0.133 (0.0038)
Proportion of queries with question	0.047 (0.0020)	0.043 (0.0047)	0.060 (0.0025)

Table 1. Descriptive statistics for all tasks and separately for successful and unsuccessful tasks. Values are means (1 std error in brackets).

Experiment #2: Online study
We used the larger data set collected online to test whether the hypothesis we formulated based on the lab study apply for a more diverse set of tasks.

Table 1 shows the descriptive statistics split by user success for the online study. Across all tasks in this study, the mean number of queries per task was 6.71 and the mean query length was 4.77 terms. Figure 1 (left) shows that the number of queries decreased steadily as the task success rate increased ($r = -0.92$, $R^2 = 0.85$ $p < 0.0001$). The abscissa shows the proportion of participants who reported success for a task (our measure of task difficulty), and the
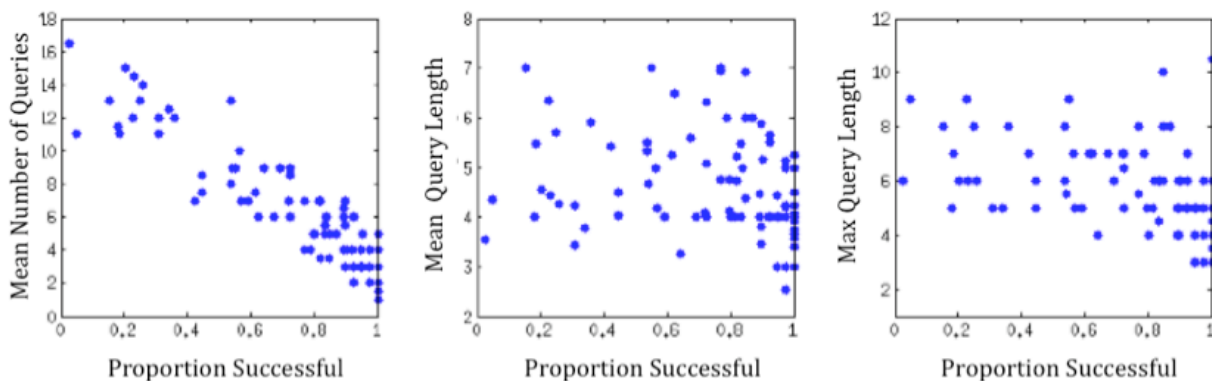

Figure 1. Graphs showing the mean number of queries, the mean query length and the maximum query length as a function of task success (the proportion of participants who were successful in each task.
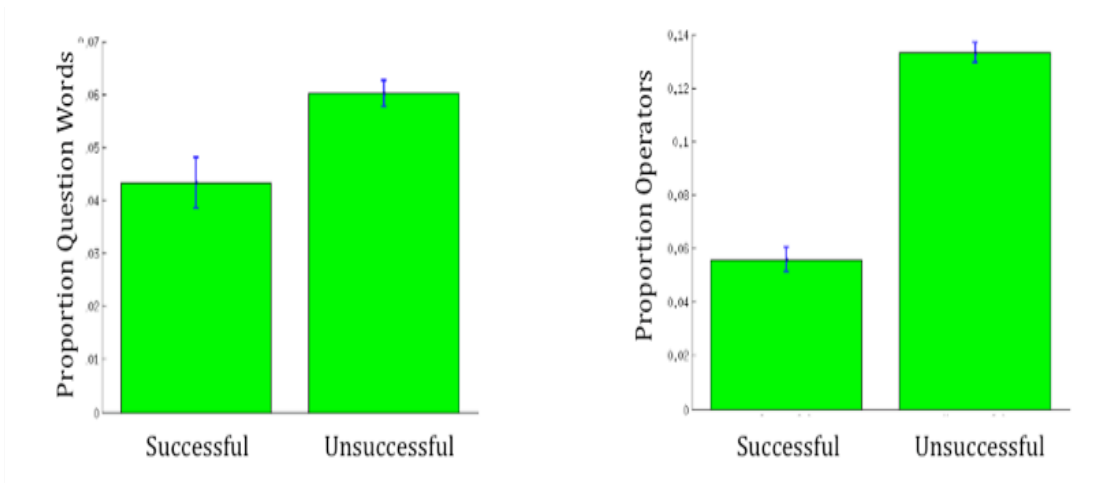
Figure 2. Graphs showing the proportion of queries with questions and the proportion of queries with advanced operators in successful and unsuccessful tasks.

ordinate shows the mean number of queries attempted. Each data point corresponds to a single task. Figure 1 (middle) shows that in addition to more queries, harder tasks (those in which fewer participants reported success) tended to have longer queries. In the tasks where more than 80% of the participants are successful, the queries tend to be between 2-5 terms long, in tasks where fewer than 50% of participants reported success, they were between 4 and 7 terms long. Figure 1 (right) shows similar data for maximum query length rather than the mean, which shows the same pattern of results. To account for the shape of the fit, we fit polynomial 2nd order regressions to both these data and showed significant relationships for both (Mean Query Length: $R^2 = 0.21$, $p < 0.0001$; Max Query Length: $R^2 = 0.29$, $p < 0.0001$).

In the lab study, we noticed that users tended to enter direct questions as queries if their keyword queries failed. To test if this hypothesis held for the larger data set, we analyzed the number of question queries for the successful and failed search tasks. In our analysis, question queries are defined as queries that start with a question word (who, what, when, how, where, why) or that end with a question mark ("?"). As shown in Figure 2 (left), although overall, the question queries were rare, users did formulate more question
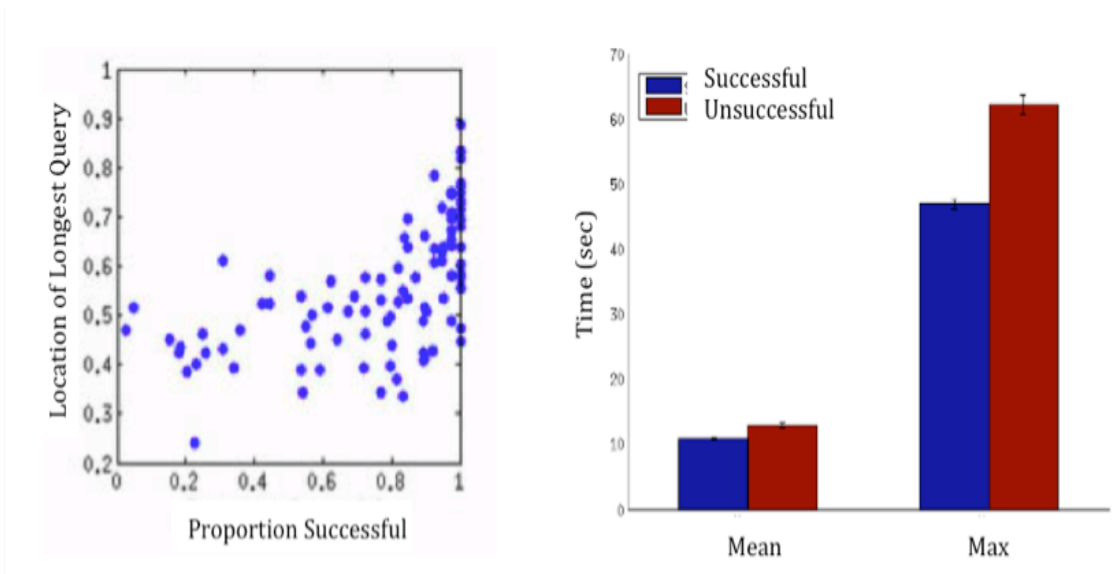

Figure 3. Graph on the left shows the location of the longest query in the search session as a function of the proportion of participants who were successful in the task. Graph on the right shows the mean and maximum time users spent on the search result page in successful and unsuccessful tasks.
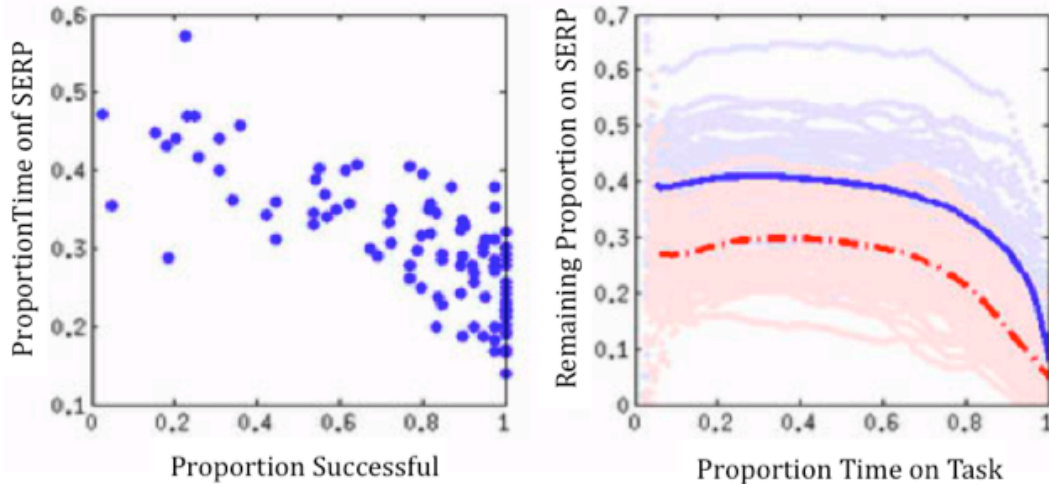
Figure 4. Graph on the left shows the proportion of total task time the users spent on the search result page as a function of task success. Graph on the right shows the proportion of remaining task time spent on the results page as a function of proportion of current task time already spent for the hardest (blue solid line) and the easiest tasks.

queries in the tasks they failed in (t(3768) = -1.972, p < 0.048).

In addition to trying the direct question approach, users seemed to try other, less intuitive strategies when the search task was difficult. Figure 2 (right) shows that the use of advanced query operators was significantly higher in unsuccessful than successful tasks (t(3768)= -6.44, p < 0.0001).

In line with the hypothesis that in easier tasks, the query refinement process often goes from a broader query towards a more specific query, Figure 3 (left) illustrates that in easier tasks, users formulate their longest query towards the end of the session (2nd degree polynomial fit: R^2 = 0.46, p < 0.0001). In the more difficult tasks the longest query tends to occur in the middle of the task, suggesting that as the usual strategy fails, they switch to other strategies that have shorter queries.

Based on the laboratory study, it seemed that when users had difficulties in the search task and they weren't sure how to proceed, they spent a lot of their time on the results page. The online data supports this hypothesis: users spent more time on the search results pages in the unsuccessful tasks as compared to the successful tasks; both when comparing the average and maximum time on the search results page (see Figure 3 right.: Mean: t(3807) = -6.91, p < 0.0001; Max: t(3807)=-10.63, p < 0.0001).).

When comparing the difficult and easy tasks in how much of the overall tasks time the user spends on the search results page (vs. other webpages) users spent a larger proportion of their total task time on the search result page in the difficult tasks (Figure 4 left: r = -0.76, R^=0.58, p < 0001). Notably, for some of the most difficult tasks, users

ended up spending almost half of the total task time on the result page. Figure 4 (right) plots the proportion of remaining task time spent on the results page as a function of proportion of current task time already spent for the hardest and the easiest tasks. . The dark blue (solid) and red (dotted) lines are the means for the hardest and easiest (median split) tasks. Light blue and red lines are individual tasks of the 2 types. Participants spent a greater proportion of time on the results page for the hard than the easy tasks and were significantly more likely to do so later in the task.

DISCUSSION

Our results showed that in unsuccessful tasks (compared with successful tasks):

- users formulated more question queries,

- they used advanced operators more often,

- they spent a longer time on the results page (both on average, and when looking at the maximum time in the search session),

- they formulate the longest query somewhere in the middle of the search session (in successful tasks, they are more likely to have the longest query towards the end of the search session), and

- they spent a larger proportion of the task time on the search results page.

When comparing search behavior we observed in our study to that reported in earlier studies, our users used longer queries, they had more queries per session, and they spent slightly longer on the search results page. One possible reason for these differences is that our study only included one task type – closed informational queries – and we

specifically included very hard tasks. We did not have any navigational or open-ended informational tasks, which are likely to be easier and as such, bring the overall query and session metrics down.

Interestingly, the overall frequency of using advanced operators was smaller in our study than that reported by others [14, 23]. In tasks where users failed, their usage of operators was comparable to the numbers reported earlier. It is possible that overall, users are now using advanced operators less frequently than they were before. This view is supported by the numbers reported by White and Morris [26], who found their log data to have only 1.12% of queries containing common advanced operators. It is plausible that since search engines seem to work well without operators [7], users have learned to type in simple queries and to rarely use complex operators unless they are really struggling.

In this study, we focused on only search task type, namely, closed informational search tasks. It is possible that some of our findings only apply to this specific task type - for example, it is hard to imagine users formulating question queries in open informational search tasks, where the goal is to simply learn something about the topic. However, it is also possible that users will be less frustrated in tasks where their goal is more open-ended. When trying to find a specific piece of information, it is obvious to users if they are succeeding or not. In less well-defined tasks, they are probably learning something along the way and frustration might not be as common. Thus, since our goal is to understand what behavioral signals could be used to determine that users are frustrated, we feel that closed tasks are at least a good place to start.

We did not specifically control for or evaluate the participants' familiarity with the search topics. Generally, domain expertise is known to affect the search strategies – users who are domain experts presumably have easier time thinking of alternative ways to refine the query should the original query fail to give satisfactory results. In our study, we had a large number of users and search tasks covering a wide variety of topics, so the differences in domain expertise are unlikely to have had a systematic effect on the results. Further research is needed to study whether users with different levels of domain expertise show different behavioral signals in successful and unsuccessful tasks.

When searchers are participating in either a laboratory or an online study, their level of motivation is different than it would be if the tasks were their own and they were trying to find the information for their personal use. Arguments can be made either way: maybe normally, in their day-to-day life, the closed informational tasks are just about random facts and if the information cannot be found easily, the user will just give up without displaying the frustrated behaviors we discovered (it was clear that in the laboratory, users clearly wanted to find the answers and became agitated when they could not find them). Whatever the general type

of user response, closed informational tasks are real tasks that people do using search engines - and although search engines are presumably good at helping people with these tasks, they will fail from time to time.

Given the varying findings related to expert strategies and how they relate to success in search tasks, it is difficult to make a direct comparison between the current findings and those of studies focusing on experts' and novices' strategies. However, some of the findings suggest that when users begin to have difficulties in search tasks, their strategies start to resemble those of novices. For example, our studies showed that when users are failing, their query refinement process becomes more unsystematic. Fields et al. [8] and Aula and Nordhausen [3] reported unsystematic refinement to be a typical strategy for novices or less successful searchers. Also the finding that users spend a longer time on the search result pages when they fail in the task resembles the behavior of less experienced searchers. Another study [2] suggested that a slower "exhaustive" evaluation style is more common with less experienced users – and this strategy seemed to be related to less successful searching.

White and Drucker [25] suggested that users are likely to use navigator type systematic behaviors in well-defined fact-finding tasks and explorer-type more variable behavior in complex sense-making tasks. Our analysis was different from that of White and Drucker and thus, it is not clear if the search trails for failed tasks would typically be classified as resembling navigators or explorers. However, our data suggests that when users are struggling in the search task – even if the task itself is a well-defined fact-finding task – their behavior becomes more varied and potentially explorer-like. Future research should focus on systematically studying the search trails and whether the search trail can provide information on whether the user is succeeding or failing in the search task.

In our analysis of the online study data, we were restricted to using the URLs and time stamps along with the ratings from our users. In the lab study, we observed other possible signals that may be related to user becoming frustrated with the search. When frustrated and unsure as to how to proceed with the task, users often scrolled up and down the result page or the landing page in a seemingly random fashion – with no clear intention to actually read the page. Another potential signal that might be related to the user becoming somewhat desperate is when they start re-visiting pages they have already visited earlier in the session. Both of these signals are potentially trackable in real time.

CONCLUSIONS AND FUTURE WORK
This study was specifically focused on measurable behavioral signals that indicate that users are struggling in search tasks – the results are an important addition to the existing body of research focusing on successful or "expert" strategies. We demonstrated how a combination of a smaller scale lab study and a larger-scale online study

complement each other. The former provided hypotheses and the latter quantitative support for the hypotheses with a more generalizable data set.

Our study showed that there are signals available online, in real time, that the user is having difficulties in at least closed informational search tasks. Our signals together with the signals related to successful and less successful search strategies discovered in earlier research could be used to build a model that would predict the user satisfaction in a search session. This model, in turn, could be used to gain a better understanding of how often users leave search engines unhappy - or how often they are frustrated and in need of help, and perhaps an intervention, at some point during the search session.

REFERENCES
1. Aula, A. (2003) Query Formulation in Web Information Search. Proc. IADIS WWW/Internet 2003, 403-410.

2. Aula, A., Majaranta, P., and Räihä, K.-J. (2005) Eye-tracking reveals the personal styles for search result evaluation. Proceedings of Human-Computer Interaction - INTERACT 2005, 1058-1061.

3. Aula, A. & Nordhausen, K. (2006) Modeling successful performance in web searching. Journal of the American Society for Information Science and Technology, 57(12), 1678-1693.

4. Aula, A. and Siirtola, H. (2005) Hundreds of folders or one ugly pile – strategies for information search and re-access. Proc. INTERACT 2005, 954-957.

5. Brand-Gruwel, S., Wopereis, I. and Vermetten, Y. (2005) Information problem solving by experts and novices: analysis of a complex cognitive skill. Computers in Human Behavior, 21, 487-508.

6. Downey, D., Dumais, S., Liebling, D., and Horvitz, E. (2008) Understanding the relationship between searchers' queries and information goals. CIKM'08, 449-458.

7. Eastman, C.M. and Jansen, B.J. (2003) Coverage, relevance, and ranking: the impact of query operators on web search engine results. ACM Transactions on Information Systems, 21(4), 383-411.

8. Fields, B., Keith, S. and Blandford, A. (2004) Designing for expert information finding strategies. Proc. of HCI 2004, 89-102.

9. Granka, L.A., Joachims, T. Gay, G. (2004) Eye-tracking analysis of user behavior in WWW search. Proc. SIGIR'04, 478-479.

10. Huntington, P., Nicholas, D., and Jamali, H.R. (2007) Employing log metrics to evaluate search behavior and success: case study BBC search engine. Journal of Information Science, 33(5), 584-597.

11. Hölscher, C. and Strube, G. (2000) Web search behavior of internet experts and newbies. Proc. WWW 2000, 337-346.

12. Jansen, B. and Spink, A. (2006) How are we searching the world wide web? A comparison of nine search engine transaction logs. Information Processing and Management, 42, 248-263.

13. Jansen, B., Spink, A. and Saracevic, T. (1998) Real life information retrieval: a study of user queries on the web. SIGIR Forum, 32(1), 5-17.

14. 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(2), 207-227.

15. Jenkins, C., Corritore, C.L. and Wiedenbeck, S. (2003) Patterns of information seeking on the web: a qualitative study of domain expertise and web expertise. IT & Society, 1(3), 64-89.

16. Kamvar, M., Kellar, M., Patel, R. and Xu, Y. (2009) Computers and iPhones and mobile phones, oh my! A logs-based comparison of search users on different devices. In Proc. WWW 2009, 801-810.

17. Khan, K. and Locatis, C. (1998) Searching through the cyberspace: the effects of link display and link density on information retrieval from hypertext on the World Wide Web. Journal of the American Society for Information Science, 49(2), 176-182.

18. Lazonder, A.W., Biemans, H.J.A. and Worpeis, I.G.J.H. (2000) Differences between novice and experienced users in searching information on the World Wide Web. Journal of American Society for Information Science, 51(6), 576-581.

19. Navarro-Prieto, R., Scaife, M. and Rogers, Y. (1999) Cognitive strategies in web searching. Proc. 5th Conference on Human Factors and the Web. Retrieved September 15, 2009, http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/

20. Rose, D.E. and Levinson, D. (2004) Understanding user goals in Web search. In Proc. WWW 2004, 13-19.

21. Silverstein, C., Henzinger, M., Marais, H. and Moricz, M. (1999) Analysis of a very large web search engine query log. SIGIR Forum, 33(1), 6-12.

22. Saito, H. and Miwa, K. (2002) A cognitive study of information seeking processes in the WWW: the effects of searcher's knowledge and experience. Proc. WISE'01, 321-333.

23. Spink, A., Wolfram, D., Jansen, B.J. and Saracevic, T. (2002) From e-sex to e-commerce: web search changes. IEEE Computer, 35(3), 107-109.

24. Teevan, J. Alvarado, C., Ackerman, M.S. and Karger, D.R. (2004) The perfect search engine is not enough: a study of orienteering behavior in directed search. Proc. CHI'2004, 415-422.

25. White, R.W. and Drucker, S.M. (2007) Investigating behavioral variability in web search. Proc. WWW 2007, 21-30.

26. White, R.W. and Morris, D. (2007) Investigating the querying and browsing behavior of advanced search engine users. Proc. SIGIR 2007, 255-262.