Slow Search: Information Retrieval without Time Constraints

Jaime Teevan[1], Kevyn Collins-Thompson[2], Ryen W. White[1], Susan T. Dumais[1] & Yubin Kim[3]

[1]Microsoft Research, Redmond, WA USA
[2]University of Michigan, Ann Arbor, MI USA
[3]Carnegie Mellon University, Pittsburgh, PA USA

{teevan,ryenw,sdumais}@microsoft.com, kevynct@umich.edu, yubink@cmu.edu

ABSTRACT

Significant time and effort has been devoted to reducing the time between query receipt and search engine response, and for good reason. Research suggests that even slightly higher retrieval latency by Web search engines can lead to dramatic decreases in users' perceptions of result quality and engagement with the search results. While users have come to expect rapid responses from search engines, recent advances in our understanding of how people find information suggest that there are scenarios where a search engine could take significantly longer than a fraction of a second to return relevant content. This raises the important question: What would search look like if search engines were not constrained by existing expectations for speed? In this paper, we explore *slow search*, a class of search where traditional speed requirements are relaxed in favor of a high quality search experience. Via large-scale log analysis and user surveys, we examine how individuals value time when searching. We confirm that speed is important, but also show that there are many search situations where result quality is more important. This highlights intriguing opportunities for search systems to support new search experiences with high quality result content that takes time to identify. Slow search has the potential to change the search experience as we know it.

Author Keywords

Slow search, crowdsourcing, speed, information retrieval.

ACM Classification Keywords

H.3.3 Information Storage and Retrieval: Information Search and Retrieval – *search process.* H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

INTRODUCTION

Although we live in a world where everything from communication to information processing to transportation is getting increasingly fast, in recent years there have been a number of "slow movements" that advocate slowing down the speed at which actions are taken in exchange for improved quality. The slow food movement is perhaps the best known such movement, with proponents encouraging traditional and diverse methods for preparing meals [32]. However, the notion has expanded to include slow parenting, slow travel, slow technology [22], and even slow science [20]. Building on these various slow movements, in this paper we discuss *slow search*, where additional time is used to provide searchers with a higher quality search experience than would be attainable given conventional strict time constraints. In this paper we focus on how search engines can make use of additional time to identify better results, but note that the search experience could also be designed to encourage people to use additional time to reflect and learn about the topic of their search [16].

To achieve near-instantaneous response times, search engines make a number of compromises to reduce computational costs, thus sacrificing potential relevance gains in favor of a rapid response to a user's query. For example, they limit the complexity of features or models used to identify relevant documents by making highly simplistic assumptions about language, such as treating text as an unordered "bag of words." The resulting fast, word-oriented matching ignores the rich semantics of text but is an efficient way to capture some of the effective similarity between queries and documents. They also limit the set of documents searched for any given query by using approaches such as search-result caching [19] and index tiering [5], even though this can mean missing relevant content, and incurring increased infrastructure costs [4].

Rapid response times have been targeted by search engines for good reason. Research suggests that when Web search engines are even slightly slower to return search results than normal, the delay leads to significant decreases in user engagement [8]. Although searchers have grown accustomed to near-instant responses to their queries, recent advances in our understanding of how people find information suggests that there are some scenarios where a search engine could take significantly longer than a fraction of a second to identify and display relevant information to users.

While it is unlikely that individuals will ever be willing to wait for a search engine to return results for a navigational query or to look up a readily-available fact, people often engage in rich, involved search tasks where time is less pressing [28]. A person planning a vacation, for example, or performing medical research following professional diagnosis may be willing to engage with the search engine differently, by, for example, waiting a significant amount of time for the best possible results. Slow search could also be useful in supporting tasks that span multiple sessions [28] or continue across multiple devices [41]. As search engines improve their ability to predict if and when a person will return to a task [28], the time between tasks could be used to monitor previously-viewed information for change or identify new information and summarize it, thus improving the user experience when the search task is eventually resumed.

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. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
HCIR '13, October 03 - 04 2013, Vancouver, BC, Canada
Copyright 2013 ACM 978-1-4503-2570-7/13/10…$15.00.
http://dx.doi.org/10.1145/2528394.2528395

This paper studies how people experience and value speed during search, and explores the viability of slow search as an alternative to current speed-focused approaches. To paint a picture of why speed has historically been crucial in the development of search engines, we begin by reviewing related literature and analyzing the logs of the Microsoft Bing search engine. We show that there are instances where speed appears to be less important than others, and discuss scenarios where search engines might be able to capitalize on additional time to provide a better search experience rather than faster results. Using two user surveys, we reveal how people trade off quality and speed. We then discuss how search algorithms could make use of extra seconds, minutes, or hours to improve the search experience, and highlight several practical considerations for slow search. The overall aim of this paper is to inspire additional research on how search experiences can be optimized when less constrained by time.

SPEED, SPEED, SPEED

The speed at which results are produced is a crucial consideration in the development of search technology. In this section we discuss why this is the case by reviewing related research on latency and search engine interaction, and by performing large-scale query log analysis to understand latency effects in a natural search setting.

Related Work on Latency and Search Engine Interaction

Intensive research and engineering efforts have been devoted to achieving low latency in large, complex computing systems such as search engines [15]. Since the early days of human computer interaction, researchers have studied the influence of system response time on the success, speed, and satisfaction of interactions [34]. In general, these studies have found that rapid responses (i.e., less than a second) are preferred and can increase user productivity.

Search engines in particular are designed to target speed. Modern web search engines deliver results rapidly because searchers interact more with them than they do with slower results, and because fast results are perceived as being of higher quality. For example, Google conducted online experiments where they intentionally injected server-side delays, ranging from 100 to 400 milliseconds, into the search results to observe changes to people's behavior. They found that increasing the load time of the result page by as little as 100 milliseconds decreased the number of searches per person. These differences increased over time and persisted even after the experiment ended and there were no longer any delays [35]. In similar experiments, Bing added server delays ranging from 50 to 2000 milliseconds. They observed decreases in queries and clicks, and an increase in time to click, with larger effects with more delay [35]. Recognizing the importance of speed to users, Google added site speed (i.e., how quickly a site responds to requests) as a relevance signal in search ranking [21]. As testament to the importance of retrieval latency to searchers, the retrieval time is also highlighted by search engines alongside other result information, e.g., "About 13,000 results (0.29 seconds)".

Speed appears to be so important that even improvements to search engines that seem like they should unambiguously impact the search experience in a positive way can lead to negative outcomes if they increase latency. For example, when Google experimented with returning 30 results instead of 10, they found that traffic and revenue in the experimental group dropped significantly. One likely explanation for why this occurred is that the additional search results required an additional half-second to load [18].

Unique Query	Navigational	Informational
Total	2274	2457
With at least one click	2065	1819

Table 1. The total number of queries in our analysis. When studying time to click, only query instances with at least one click were included. The unique queries were then filtered to ensure sufficient data.

Query Log Analysis of Speed and Interaction

However, user preference for speed can be nuanced. We extend the above findings using the logs of the Microsoft Bing Web search engine. We discover that user engagement does not always vary linearly with time and that the effects depend on query type.

Methodology

Unlike previous studies that inserted fixed delays into the load time of search-result pages, we study naturally-occurring variation in search result latency. Such delays are representative of a user's typical search engine usage experience. Using the Bing query logs from users in the United States English language locale, we examine the post-query behavior for the week of January 5, 2013 to January 11, 2013 for queries with different page load times.

Query behavior can vary significantly by query, with factors such as the task and result quality influencing the number of results clicked and the speed with which they are clicked. Thus it seems likely that a delay in search engine response time would impact different tasks in different ways. For this reason, in our analysis we only compare behavior within queries. User behavior following a particular query with a fast page load time is only compared to instances where that same query resulted in a slow page load time.

For each query, we categorized the query instances into five different groups based on the page load time. The fastest group we studied contained instances of load times between 500 and 700 milliseconds. Each subsequent group ranged 200 milliseconds (e.g., 700 to 900 milliseconds) through the slowest group studied, which contained query instances that took from 1300 to 1500 milliseconds to load. Outliers falling outside of these ranges were not used in the analysis. We then selected the 4731 unique queries with at least 100 instances in each group for our analysis.

To further understand how query task impacts behavior, we look at the impact that search engine response time has for navigational and informational query types. *Navigational* queries, such as [*facebook*] targeted specific web pages (facebook.com in this case). In contrast, *Informational* queries such as [*state abbreviations*], are intended to find information about a topic or answer a question. The navigational and informational queries in our data were identified using a proprietary classifier from the Bing search engine.

The two post-query behaviors that we consider are *abandonment rate* (i.e., the fraction of times the search result page is loaded and there is no click) and *time to first click* (i.e., the time between the page loading and the first result click). Because a click must be present to calculate the time to first click, for this metric is only calculated for the subset of query instances where a click was present. Queries that do not have at least one click for at least 100 instances in each page load time group are not considered. Table 1 displays the number of unique queries used for each category and behavior studied.

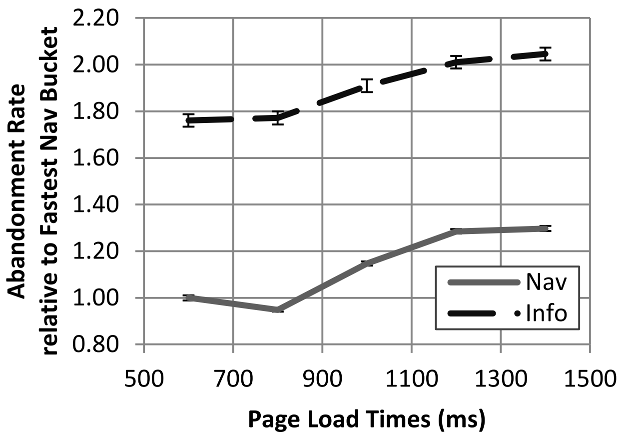

Figure 1. Changes in abandonment rate with increasing page load times for navigational and informational queries including standard error bars. In the case of navigational queries, the standard error bars are too small to be seen. Shown relative to the fastest loading navigational query.

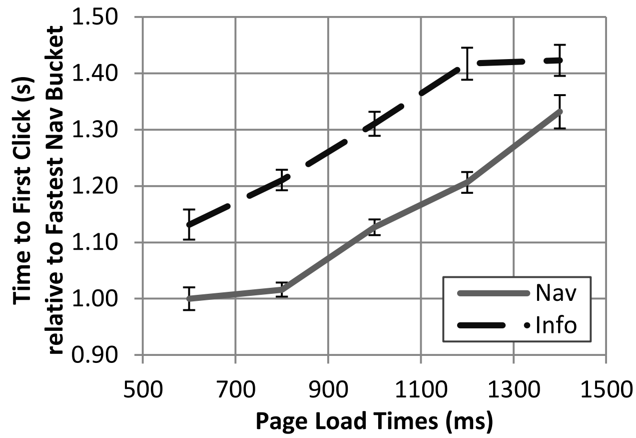

Figure 2. Changes in the time to first click with increasing page load times for navigational and informational queries including standard error bars. Shown relative to the fastest loading navigational query.

Results

Figures 1 and 2 provide summaries of people's post-query behavior as a function of how quickly the search result page loaded, macro-averaged across unique queries, with standard error bars. As expected given the findings from related studies, engagement with the search result page decreased as its load time increased. Both the abandonment rate and the time to click increased significantly from the fastest page load times to the slowest page load times. As shown in Figure 1, as page load increased from the quickest to slowest times, search abandonment rate increased by 29% for navigational queries and 16% for informational queries. Additionally, as shown in Figure 2, we observe that searchers were able to identify relevant content faster within result pages that loaded quickly. The average time it took a user to click on a result increased 33% for navigational queries and 25% for informational queries as the page loaded more slowly.

However, the change in user engagement did not appear entirely linear with respect to load time. In the case of abandonment, we observe a tapering effect as page load time increases for both query types. While there are increases in abandonment rate between the 700-900 millisecond range to the 1100-1300 millisecond range, the difference in abandonment rate between the 1100-1300 millisecond group and 1300-1500 millisecond group is not significant for both navigational and informational query types. This suggests that the negative impact of page load on abandonment times may taper, with longer wait times not leading to comparable increases in abandonment. When looking at the time to first click we observe a similar tapering effect, but this time only for informational queries. The difference between the 1100-1300 millisecond group and 1300-1500 millisecond group is not significant for informational queries. In contrast, for navigational queries the difference is significant and the growth appears roughly linear.

Our analysis of the log data suggests that although speed impacts user experience, the effect is non-linear and can impact different query types in different ways. This suggests that there may be some tasks for which search engines could more easily trade off speed with better result quality than others.

NOT ALL SEARCHES NEED TO BE FAST

The fact that search engine speed impacts navigational queries differently than informational queries is just one example of how a nuanced interpretation of time may be more appropriate than a pure quest for speed in all search circumstances. This is consistent with related research in graphics interaction that has shown that rapid response times can sometimes be detrimental. Barber and Lucas [5] found that when system response times were too fast errors increased because people responded to the system too quickly. In this section we show how existing research reveals that there are a number of information seeking scenarios where results quality may be more important than result speed.

Long-Term Tasks and Exploratory Search

One thing that suggests that people can make do with search results that take longer than a fraction of a second is that in practice searchers do not often actually find what they are looking for after just one query. Over half of the time that people spend using a search engine they are engaged in multi-query search sessions that take minutes or hours [17]. Queries are often not issued in isolation, but rather as part of a larger information seeking task. In such cases, what is being sought can often be much more complex than what is captured in a single query and set of search results. Although search engines currently perform well at ad hoc retrieval with simple queries, they do not support more complex or exploratory tasks, which may span multiple queries or sessions, as effectively [44].

Recent research is paving the way toward support for complex tasks over time. Tyler et al. [38] show that people regularly return to previously issued search queries. For these involved tasks users may know that they would be willing to wait for high quality information. People could actively request slow search or a search engine could suggest that it be employed when a user has been iterating or floundering on a topic for a while or seems likely to switch to another search strategy or even another search engine [42]. Reconnaissance agents, e.g., [26][30], have been proposed as a way in which search engines could help searchers by finding relevant information in the background as they engage in other tasks, search-related or otherwise. System support could also be provided to help people make decisions on which search engine to use based on performance estimates for the current query [43].

By making informed predictions about what people will search for, slow search tools could also utilize additional time to produce high-quality search results without requiring people to wait. Kotov et al. [28] found that it is possible to predict when someone will return to a task at a later date. Search sessions can also be predicted across devices. For example, a person might search for a restaurant on their desktop computer, and then look up directions and menu information later from their mobile phone as they head to the restaurant. Wang et al. [41] found that they could predict task resumption on a target device using behavioral, topical, geospatial, and temporal features. Techniques like this could be used to determine what information a user is likely to seek in the near future, enabling a search engine to proactively—and slowly—find relevant content.

Social Search and Question Asking

Different expectations for speed during search can also be found by studying people's information seeking behaviors outside of search engines. When speed is not a primary factor, people often ask questions of others through email, web forums, or social networking sites. These conversations tend to be asynchronous, with little or no expectation of an instantaneous response. Search tools could be used to augment these conversations. Within a social context there is an opportunity for a search engine to take more time and devote more resources to the question than can be done in the few milliseconds typically allotted following a web search query.

Response times in online question-and-answer forums tend to take on the order of hours. For example, Zhang et al. [45] reported that when expert Java users posted questions to the Java Developer Forum, the average time to receive a response was nearly 9 hours. Hsieh and Counts [24] reported that the average time to receive an answer to a question posted to Microsoft's Live QnA site was almost 3 hours. Likewise, responses to questions posted to social networks appear to be slow. Morris et al. [31] found that 24.3% of the people they studied received a response to a question posted on Facebook or Twitter in 30 minutes or less, 42.8% in one hour or less, and 90.1% within one day.

Because the norms involved with question asking are different from those associated with search engine use, people appear to be willing to wait a significant amount of time for responses that come from people. For example, even though Morris et al. [31] found that most people received responses to their social network questions after an hour or more, 93.5% reported feeling like their questions were answered promptly.

The trade-offs information seekers make between time and quality are somewhat clearer in social contexts than they are for more traditional search engines. In these contexts they can affect both question asking and question answering. Aperjis et al. [2] studied Yahoo! Answers to explore how question askers tradeoff the value of the answers they received with the time that they spent waiting for a response. They found that askers would wait longer to receive additional answers when they have received only a few responses. In addition, the speed with which questions are answered is important. Anderson et al. [1] found that the faster an answer is received on the StackOverflow site, the more likely that answer was to be chosen as the best one.

Technological Limitations to Speed

Slow search approaches could be especially valuable when people have limited access to the Internet, with their connection intermittent, slow, or expensive. For example, in the developing world some areas have access to the Internet for only a limited amount of time each day or each week (e.g., a traveling Wi-Fi bus may briefly visit a village). In such cases it can be difficult for people to employ traditional search strategies, such as rapidly iterating on a query.

People can strategically adjust their Web search behavior to handle these situations. Chen et al. [10] looked at how Web search and browsing behavior differs when connections are slow, and found that people employ strategies to address the challenges that arise. A better solution may be for search systems to be explicitly designed to address such challenges. For example, Thies et al. [39] describe an email based search engine where users in low-connectivity communities can issue queries and receive responses via email. The query and result set are processed in minutes or hours, optimizing for bandwidth rather than response time.

Mobile phones can also have limited connectivity, and slower search processing times may be acceptable when most of the latency a searcher observes comes from network latencies in fetching data to the device. For example, Jisiklog is a Korean question answering service for mobile users [29]. Users submit questions via SMS and receive responses generated using crowdsourcing in a matter of minutes rather than seconds. Despite the wait, people pay for this information because of its high quality.

Likewise, future space travelers may also appreciate slow search. It takes over 25 minutes for information to travel from Mars to Earth and back again. If a search engine were to take an additional few minutes to identify better results during the round trip, it is likely the searcher would not even notice the extra time invested, but would be likely to appreciate the enhanced result quality.

SURVEYING ATTITUDES TOWARDS TIME IN SEARCH

To better understand people's attitudes towards waiting for search results, we conducted two user surveys. These surveys provide us with examples of search tasks with a variety of time constraints, and through respondents' subjective perceptions of how slow search could operate in practice, they provide us with insight on how participants might trade off quality and speed.

Methodology

The two surveys we conducted were designed to elicit input about people's time preferences with respect to search at varying levels of depth across two different populations. In both surveys we asked participants to tell us about the most recent search they had conducted on a web search engine. The first survey, which we refer to as the *detailed survey*, also includes free text answers from 141 volunteers at a large technology company (Microsoft). The second, which we refer to as the *quality survey*, looks at more levels of quality over a larger population, with responses collected from 1335 crowd workers, but includes less detail per response.

Detailed Survey

In the detailed survey we asked participants to tell us about the most recent search they had conducted on a Web search engine, reviewing their search history if necessary to remember. To avoid navigational queries, we requested that the search involve more than one query. Specifically, the question asked,

"Briefly describe the last thing you searched for using a web search engine that involved more than one query. If you need help remembering, click the Bing or Google link [links provided] to see your past queries."

Participants provided a free-text description of the search, such as "team morale event ideas" or "what happens if name in airline

		n	Searching	Waiting
Total		141	10	5
Task importance	Very	83	15	3
	Less	58	5	5
Result quality after searching	Best	20	3	2.5
	Ok	89	10	3
	Bad	32	30	10
Task urgency	Urgent	48	5.5	4.5
	End time	55	15	5
	None	38	8	3.5

Table 2. The median amount of time (in minutes) that participants in the detailed survey said they spent actively searching or were willing to wait for results under different conditions.

ticket has a typo error." They also reported how important the search was to them (very important, important, somewhat important, or not important), and judged the overall quality of the results they found (best possible, acceptable, or not acceptable).

To understand participants time preferences with respect to their search, we asked them to share whether the task was time sensitive (needed urgently, by a particular time, or anytime) and how long they spent searching (in minutes). We also wanted to build a picture of how long they might be willing to wait for the search engine to identify a high quality response to their search. To do this, participants were asked to "imagine that you could have found comparable information just by waiting for a response to your first query, rather than by issuing multiple queries and actively sifting through the results. While you were waiting, you could choose to do other things." We then asked them to report, in minutes,

"How long would you be willing to wait if you knew the search engine would identify the best possible response without any additional effort on your part?" as well as "an acceptable response."

We followed this with a request for a free text response about what might make them willing to wait longer.

The survey also collected basic demographic information, including age, gender, and familiarity with search. Invitations to complete the survey were sent via email to a set of randomly selected people from within Microsoft Corporation, and 141 responses were received. Consistent with employee demographics, most respondents were male (110 or 78.0%) and in their 30s and 40s (115 or 81.6%). Most respondents were very familiar with technology and Web search, with 105 (74.5%) holding a technical position and all but three (2.2%) reporting that they searched the Web at least daily.

Quality Survey

While the data collected from the detailed survey provides a rich picture of user time preferences with respect to Web search, the detail requested from participants limited the number of responses we could collect, and respondents were limited to a population of people working at a technology company. This may introduce bias into the survey responses and we wanted to address the potential for that to occur. To more accurately model the tradeoff between quality and time, we also conducted a survey to collect data at a larger scale for more levels of potential result quality improvement from a large number of crowd workers.

The quality survey contained only four simple multiple-choice questions. Each question asked,

"Think about the last time you used a search engine and had problems finding something. If you could use a new search engine that took more time to find better answers, while you were free to go and do other things, what's the longest you'd be willing to wait for a [quality-level] answer?"

Four different quality levels were used: "A perfect answer", "A much better answer", "A better answer", and "A slightly better answer". Participants were given five answer options, based on what appeared to be meaningful wait durations from the data collected in the detailed survey: 1 minute, 15 minutes, 60 minutes, 360 minutes, or longer.

Rather than recruiting from a single institution as we did with the detailed survey, participants were recruited via TellWut.com, an online survey platform that collects crowdsourced responses. Demographic information about the respondents was not available. The survey received 1335 responses.

Results

We now examine what the two surveys reveal about the time people spend searching, how willing they are to wait for results, and how they trade time off with quality in their searches.

The topics that participants reported having recently searched on in the detailed survey were rich and varied. For example, one participant wanted to learn how to recover data from a flash drive, and another wanted to find the name of an artist whose work he had seen recently. Eighty three (58.9%) participants said their search task was "very important," while only two said it was "unimportant." Unfortunately, however, many of the respondents reported that Web search engines were unable to adequately satisfy their information needs. Only 20 (14.2%) reported that they found the best possible results by searching, while 32 (22.7%) said that the results they obtained were unacceptable.

For 48 (34.0%) of the searches reported in the detailed survey, the information being sought was needed urgently (e.g., one participant wanted to find a live online broadcast of the 2013 US presidential inauguration). Another 55 (39.0%) required results by a particular time (e.g., tax forms), and 38 (27.0%) had no time constraints (e.g., someone wanted to find the number of homeschooled children in the US). In general, it seems that the more urgent the information need, the more important the search task was to that individual; 68.8% of those who said they needed the information urgently said the task was "very important," compared with 61.8% of those who needed it by a particular time, and 42.1% of those who had no time constraints.

Time Spent Searching

Because participants in the detailed survey estimated the amount of time they spent to complete their search task, the survey is useful for understanding the duration of their search experience as a function of task importance, result quality, and task urgency. These results are reported in the column titled "Searching" in Table 2. Overall participants reported devoting a median of 10 minutes to searching. Note that while the table reports medians, the actual amount of time different individuals reported searching varied widely. Twenty five respondents (17.7%), for example, said they searched for over an hour. Only 17 (12.1%) respondents said they spent the least amount of time possible (1 minute),

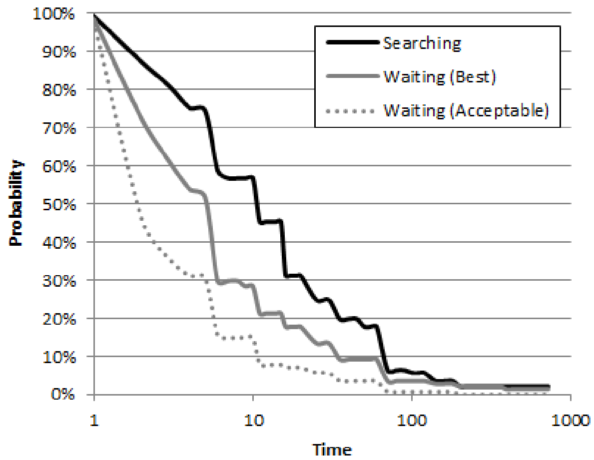

Figure 3. The probability that survey participants searched or were willing to wait at least *T* minutes for their search results. (Time on log scale.)

suggesting that for close to 90% of respondents, the search engine could have used additional on-task time to support their information searching.

We represent the variation in how long people reported searching graphically in Figure 3 with the solid black line. The figure represents time (in minutes) along the *x*-axis, and the probability that a participant reported searching for at least *T* minutes along the *y*-axis. We call this curve an *impatience curve* because it represents how impatient people are likely to be in finding what they are looking for. (The gray lines represent how long people reported being willing to wait for results, and are described in the next section.) There are close connections here to the time-based gain work of Smucker and Clarke [36, 37] who also considered that gains arrive at successive points in time. They modeled the survival probability that a user would continue a search process to time *t* with a decay function *D*(*t*), which corresponds more or less to our impatience curves, with differences in the nature of the specific process modeled (e.g., traversing a ranked list versus waiting for search results).

Not surprisingly, when participants believed that their search tasks were important, they reported devoting significantly more time to the search process. Table 2 reports a median 15 minutes spent on very important tasks, versus only 5 minutes when the task was less important. Participants also reported spending a lot more time searching when they could not find what they were looking for. They reported finding high quality results quickly (median 3 minutes), while it took 30 minutes before giving up after having only found low quality results. Search tools could offer additional support that provides better quality results at the expense of time for these long, difficult sessions.

Participants reported spending less time searching (median 5.5 minutes) when the task was more urgent than typical, but they also reported spending less time searching (median 8 minutes) when there was no time urgency in finding the results. They reported spending the most time searching (median 15 minutes) for content that needed to be found within a particular time frame. These search tasks were reported to be as important as the urgent tasks, but they were reported to be twice as likely to yield poor results; 30.1% of the tasks that had deadline turned up "unacceptable" results, while only 18.6% of the urgent tasks did. It may be search tasks with a deadline are particularly difficult, important tasks. For these tasks, searchers should be able to specify their time constraints in advance because they are clearly defined in the context of the search.

Willingness to Wait for Results

In addition to telling us how long they actually spent searching, participants in the detailed study estimated the amount of time they would "be willing to wait if [they] knew the search engine would identify the best possible response without any additional effort on [their] part." These results are shown in the second column in Table 2 (titled "Waiting") and as the gray impatience curves in Figure 3.

The solid grey line in Figure 3 represents the probability that a participant reported being willing to wait at least *T* minutes for the best possible search results. Like the curve of time spent searching (black line), its shape is roughly exponential. Despite the fact that the question explicitly stated that, "while you were waiting, you could choose to do other things," participants still reported devoting more time to actively searching for (mostly just acceptable) results than they said they were willing to wait for the best possible results. The median amount of time that they reported being willing to wait (five minutes) was half the median amount of time they reported spending searching,

Only 36 (25.5%) participants could imagine waiting for the best possible results longer than they actively searched. They said that they were willing to wait tens, or even hundreds, of minutes for the search engine to identify high quality results. The remaining 86 (61.0%) of participants had difficulty envisioning a search engine that would sacrifice speed for quality. When these people were asked what would make them consider waiting, they tended to reply that they would like to "see fast results always," a sentiment explicitly echoed by 39 (27.7%) of all participants. Despite being unwilling to wait, there was some indication that these participants would wait in different circumstances. For example, one participant who reported spending 30 minutes searching for details of the Visual Basic compiler said,

"After finishing ... without a useful answer, I sent off an email and was willing to wait much longer, as the result I would get from the experts would be definitive."

As this quotation suggests, the reason that people were unwilling to wait seems to relate primarily to the fact that they did not trust the search engine to identify better information than they would find by actively searching themselves. For example, one participant said that he would be willing to wait longer, "if [he] knew that the results would give [him] exactly what [he] needed," and this sentiment was brought up in the comments by 58 (41.1%) participants. The more important the search task was the less participants said they were willing to trust the search engine to proceed without active intervention. Table 2 shows that although people said they spent three times as much time searching on very important tasks (15 minutes) than less important tasks (5 minutes), they said they were less willing to wait (3 minutes versus 5 minutes).

When people were unable to find what they were looking for using the search engine, they became more willing to try alternate approaches. As one participant stated, "Usually, I can either find the answer I want in two seconds, or I can't find it at all. So, for the 'can't find it at all' questions, I would be willing to wait essentially any amount of time." As shown in Table 2, participants who reported finding poor results while searching were willing to

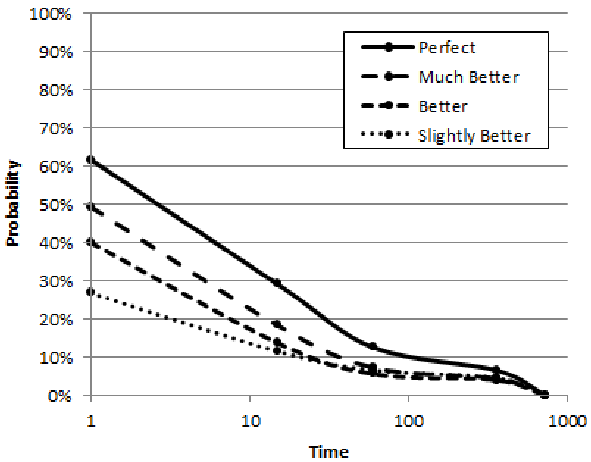

Figure 4. The probability participants were willing to wait at least *T* minutes for their search results for different answer quality levels. (Time on log scale.)

wait four times as long for the search engine to identify content as people who find high quality results (10 minutes versus 2.5 minutes).

Trading Off Time with Quality

To understand how people trade off time with quality, we must study the length of time that person might be willing to wait for results of different quality levels. For this reason, in addition to asking participants how long they might be willing to wait if they knew the search engine would identify the best possible response for them without any additional effort, we also asked about how long they might be willing to wait if they knew the search engine would only identify an acceptable response. The median length of time reported was, not surprisingly, significantly lower in this case, at only one minute. Figure 3 shows the impatience curve for acceptable responses (dashed gray line) compared with the impatience curve for the best possible responses (gray line).

Because the tradeoff between quality and time needs to be well understood to accurately model it, we constructed impatience curves at a larger scale for more levels of potential result quality improvement using the quality survey. Figure 4 shows the impatience curves for the four quality levels. Not surprisingly, people reported being most willing to wait for a perfect answer, and least willing to wait for an answer that is only slightly better than what they identified by searching on their own. Although the differences among most quality levels decrease as the amount of time increases, a perfect answer appears to be something that people are much more willing to wait a long time for.

There is greater variation at the one-minute level for the curves created from the quality survey than observed in the detailed survey. This may be because the quality survey was designed to better represent people who might be unwilling to wait at all by asking *whether* they would wait a minute rather than *how many* minutes they would wait. Additionally, participants in the quality survey were asked merely to think about their last search, and not their last search that contained at least two queries, as was the case in the detailed survey.

Summary

In this section we have offered insight on searchers' perceptions of the tradeoffs between time and result quality. We showed that people often pursue important tasks and that since these tasks take time (many last over one minute), there is opportunity to help searchers in close to 90% of these tasks. We showed that searchers who obtain poor quality results are more willing to wait than those with high quality results, and those seeking a perfect answer are more willing to wait. Tasks where there were specific time constraints (e.g., information needed within next 30 minutes) were important were equally urgent but more likely to yield poor results than typical urgent queries; users could specify these constraints to the system in advance. Beyond answer quality we showed that people may be unwilling to wait because they do not trust that the search engine would be able to understand their needs and find better results given more time. Indeed, over half of participants had difficulty envisioning a search scenario where they would not want results instantly, even though they actually spend significant time searching and waiting on responses to questions posed using other means (e.g., email). To realize real user value from the extra time search engines may have for some search tasks, we need slow search algorithms that operate under temporal constraints.

SLOW SEARCH ALGORITHMS

In this section we discuss related research that has considered the tradeoff between time and result quality in search results. We discuss ways that slow search techniques can be used to improve search quality over the course of seconds, minutes, or longer.

Improvements by the Second

With additional seconds to invest in a search, a few of the existing restrictions typically used by search engines to maximize speed can be relaxed to take just a little more time to identify better results. For example, multiple queries related to the user's query can be issued to broaden the set of candidate documents to cover different aspects of the query, to confirm the importance of key concepts in the query, or to find related terms that may be used to augment the initial query for better matching. These methods are advantageous in that they fit with the current search paradigm (meeting searcher expectations for how a search engine should operate), align with search engine infrastructure (process received query and respond near-instantly with a ranked list of query-relevant results), and are not dependent on access to resources (human or otherwise) outside of the engine. Search engines could quickly integrate such methods and compare alternatives at scale based on behavioral data.

Researchers have explored retrieval algorithms for difficult queries that trade off the computational cost of using many queries for the benefits of finding results that reflect all the relevant aspects of an information need. Crabtree and colleagues [14] were able to improve precision by 15-20% over a commercial search engine on several small sets of difficult queries, with only few queries experiencing degraded performance. The authors estimated that for typical Web searches, an average of 56 additional queries would need to be issued by their method. For the most part, these could be issued in parallel, and the authors discuss further optimizations for efficiency. However, some queries could be dependent on previous queries, necessitating additional processing time.

Improvements by the Minute

With minutes to invest, search engines can start employing entirely new approaches, such as including humans directly in the process of finding results or composing answers. Crowd-based ranking methods use human judgments to identify the most relevant existing content for a query, or potentially to generate new relevant content. Jisiklog is a Korean question answering service in which mobile searchers submit questions via SMS and receive responses generated using crowdsourcing. Responses take minutes rather than seconds, but despite the wait, people are willing to pay for this service because of the response quality [29]. Chen et al [11] propose a model for ranking that uses pairwise comparisons collected via crowdsourcing. While they use this information for offline evaluation, similar approaches could be used for ranking.

Longer Improvements

Additional time could be used not only to identify more relevant results, but also to summarize or organize the identified results, creating new search artifacts for consuming and archiving, and enabling new search experiences. Slow search applications could include monitoring previously-viewed information for changes, supporting the resumption of ongoing tasks, summarizing sets of results, finding the right collection of background material to help answer a question or learn about a topic, or finding easier alternatives to more complex content.

For example, while search engines have long connected people to documents, they are increasingly connecting people directly to information gleaned from external sources. For popular topics such as weather, movies, and definitions, search engines sometimes add custom content (e.g., "77°F, partly cloudy"). This content, known as *answers* [12], allow searchers to satisfy their information need without clicking through to a search result. In the detailed survey described earlier in the paper, respondents offered other suggestions for how additional time might be used by the search engine to create more useful results beyond a simple ranked list, including the creation of summaries, the identification of old or outdated content, the retrieval of deep Web content from pages typically not indexed by search engines, the verification of different information sources, and the translation of search results into relevant languages. Two people even expressed a willingness to wait for the search engine to identify tangential or serendipitous content.

Bernstein et al. [6] used the crowd to generate answers for queries by identifying answer candidates using aggregate search and browsing patterns, filtering those answer candidates to ones which represent directly answerable needs, using search logs and paid crowdsourcing, and extracting the answer content from the web, using paid crowds to copy and paste content from the page, then author and edit the final answer text. Creating answers using this process took on the order of 30 minutes, although recent work on real-time crowdsourcing could significantly speed up the process [7]. Crowd-generated answers have been shown to be high quality or only contain minor errors 86.7% of the time [6]. When correct, they significantly improve the user experience, particularly when the search result quality is low. While answers improved the user experience over typical high-quality results by 23.4%, they improved the user experience by 70.7% over low-quality results.

Online question asking is another type of crowd-powered search [1, 2, 29, 31] that produces answers rather than a list of results. Jeong et al. [25] employed crowd workers to create replies to people's questions on Twitter, and comment on replies from friends. Bigham et al. [7] use crowdsourcing to find answers to image-based queries. Hecht et al. [23] developed a system called SearchBuddies that responds to Facebook status message questions with algorithmic search results alongside other, friend-generated replies. For example, when a person using SearchBuddies posted a question to their network about cell phone roaming charges in Hawaii, the system replied with a link to a page titled, "do cell phones work in Hawaii," and the names of several friends who live in the state. The authors note that there are unique challenges with creating short conversational responses rather than a list of results, including determining when to respond.

Given current approaches such as those described in this section, providing the aforementioned functionalities requires a significant amount of time, and intermediate short-term results that would be valuable to the user tend not to be available in the meantime. This makes them interesting to study in the context of slow search.

Trading Off Time with Quality

Slow-search systems require computational mechanisms to determine the appropriate method given the time and resource constraints of a particular query. Effectiveness and efficiency are important topics in information retrieval, but there are not many studies that explicitly examine tradeoffs between the two. The Terabyte Track at the annual Text Retrieval Conference (TREC) was designed in part to examine efficiency issues. Büttcher et al. compared the best run and fastest run for each group [9]. Although across group comparisons are difficult, they observed that when systems spent more time processing a query because the used more complex features, the effectiveness (P@20) of the systems increased. For example, in the University of Waterloo runs, as more terms per document were added to the index, latency increased from 13 to 32 milliseconds per query but accuracy increased (0.41 to 0.48). In the University of Melbourne runs, using positional information increased latency from 55 to 229 milliseconds but accuracy also increased (0.49 to 0.51).

In the context of question answering, Azari et al. [3] studied how the type and number of query rewrites influenced answer quality. They developed Bayesian models to enable cost-benefit analyses, trading off the expected gain in accuracy of an answer with the cost of submitting additional queries. Researchers have studied speed and effectiveness tradeoffs in distributed information retrieval via the number of queries used to sample and build up an accurate picture of a particular collection. For example, Cetintas and Si [13] used utility-based optimization to automatically decide whether to download more documents for training. Researchers have also explored employing different, more intensive algorithms for hard queries [14,33]. There is also existing work on cascade models for time-sensitive ranking that has the goal of degrading effectiveness gracefully under time constraints [40].

Practical Considerations

Many of the slow search approaches discussed above also involve significant cost, since they require significant additional resources to identify relevant results. For example, the human crowd workers used in some approaches discussed earlier would likely need to be compensated for the efforts. An appropriate business model is needed in order for human-powered solutions to be cost effective at scale in search engines.

While devoting significant additional resources to create a better search experience can be expensive, slow search also offers cost-saving opportunities for search providers. Search engines currently experience significant resource demand during periods of high activity, and must develop infrastructure to handle peak loads. If some query processing were unconstrained by time, this load could be distributed more evenly to leverage underutilized resources.

Additionally, although many slow search results are likely to be highly contextual, some of the information created during slow search may be useful to other searchers. For example, online question asking websites are intended not only to meet individual's immediate needs, but also to create a database of questions with high quality answers. Slow search could be considered a type of *on-demand retrieval*, where the work to identify good results is only done when someone first needs them, but then it can be done quickly for others with the same need in the future.

SLOW SEARCH USER EXPERIENCE

Slow search will also change the user experience. Although most search engines currently represent a searcher's need using a single query, users may wish to express their slow needs more richly. For example, in online social situations people typically provide long natural language explanations of what they are looking for [31]. Rather than typing the query [*vegetarian recipe*], people provide context and detail that can be leveraged by slow search algorithms, asking questions such as, "Can anyone recommend a good spicy vegetarian recipe without tofu or mushrooms?"

In the context of a search engine it can be difficult for searchers accustomed to issuing short queries to provide additional information about their information need. However, research has shown that asking searchers to describe the context of their information needs can be valuable [27]. In addition to topical content, searchers may also want to specify other properties of the results or the slow search process itself, including timing constraints. Additional context could be identified implicitly, as is often done with personalization and contextualization. For example, if a search engine is going to work to identify new results before a task is resumed, the queries a user issues in the first session could be used to represent the searcher's slow search need.

Slow search can change how people interact with search results. Searchers are no longer engaged in a low latency dialog with the search system, but instead must wait for results to be found. Rather than waiting until a large number of relevant results have been obtained, the search engine could update the user as results are found, enabling searchers to provide early feedback about the system progress. In the survey described earlier we observed that many searchers are unwilling to trust search engines to make good use of any additional time afforded to them, even when they are looking for more complex information than a single query and set of results could capture. Search engine support is needed to help searchers intervene early and build trust in slow operations.

As part of this, slow search engines should clearly communicate the status of the slow searches and help searchers to understand the benefit of a delayed system response. These systems should also provide ways for searchers to interrupt a slow search and to easily resume a suspended search task. People in the survey suggested proposed solutions such as progress reports, dynamic result lists that visibly improve over time, and trivia in the topic area they are searching (as a way to make the wait more tolerable). Although many of the ideas suggested by participants suggest strategies for supporting active waiting, slow searchers may want to do other things while a slow search is happening. In such cases, they may want to be alerted when new content is available. For example, one participant said they would wait longer, "if there was a nice way to be notified when the result was available, that way I'm not really waiting." Notification can occur via email, text message, or other mechanisms, or even via an alert when the user returns to the search engine (e.g., on the engine's homepage). Slow search content could also be inserted into a typical search result list when a user issues a query related to a previously-issued slow search. Blending the past and the present in this way allows for the seamless integration of slow search into existing search solutions. Although there are many possible slow-search applications, little has been done to determine how to best reinstate the required context when a user returns to a slow search task.

An important goal of slow search is to free searchers from the low level processes of searching, allowing them to focus instead on task completion. While people could use the time a slow search engine spends searching on their behalf to perform other tasks, they could also use it to reflect and learn about the topic of their search. Dörk and colleagues [16] suggest slowing down the search experience by encouraging people to view result content at different levels and deviate off-topic during the course of a search. With additional time, search engines could create artifacts that do not just help users answer a targeted information need, but also help them comprehend the context of that information and learn what is necessary to fully understand it. It could be that adding additional time to the search process will not only allow search engines to identify and return the most relevant content but also enable searchers to get the most possible from the search experience.

CONCLUSION

We have explored what search might look like if the current compromises made by search systems for speed were relaxed. Through search engine log analysis and a detailed review of related work, we showed that the compromises that have historically been made to save time exist for good reason, with research suggesting that even slightly slower retrieval can lead to a dramatic drop in the perceived quality of results. However, we argue that while speed is often important, there are many information seeking scenarios where it is not, and use two large-scale user surveys to model how long people are willing to wait for results as a function of quality. In these cases, in return for increased response time, slow search algorithms may lead to greater utility for complex or difficult information seeking tasks, as well as being a better fit for scenarios involving limited-bandwidth or low-connectivity environments. We discussed several approaches that use extra time to involve the crowd or additional computational resources to provide a better search experience than might be possible within stringent time constraints. Future work includes developing slow search systems and measuring their effectiveness in different task contexts, as well better understanding searchers' willingness to wait when faced with such decisions in-situ (rather than speculatively as we with the surveys presented here). Our hope is that this work will inspire additional research into how the search experience can be improved using a more nuanced notion of time constraints.

REFERENCES

1. Anderson, A., Huttenlocher, D., Kleinberg, J. and Leskovec, J. Discovering the value from community activity on focused question answering sites: A case study of Stack Overflow. In *Proceedings of KDD 2012*.
2. Aperjis, C., Huberman, B.A. and Wu, F. Human speed-accuracy tradeoffs in search. January 11, 2010. http://bit.ly/9bTehC
3. Azari, D., Horvitz, E., Dumais, S.T. and Brill, E. Actions, answers, and uncertainty: A decision-making perspective on web-based question answering. *IP&M*, 40(5), 2004.
4. Baeza-Yates, R., Murdock, V. and Hauff, C. Efficiency trade-offs in two-tier web search systems. In *Proceedings of SIGIR 2009*.

5. Barber, R.E. and Lucas, H.C. System response time, operator productivity and job satisfaction. *CACM*, 26(11), 1983.
6. Bernstein, M., Teevan, J., Dumais, S.T., Liebling, D. and Horvitz, E. Direct answers for search queries in the long tail. In *Proceedings of CHI 2012*.
7. Bigham, J., Jayant, C., Ji, H., Little, G., Miller, A., Miller, R., Tatarowicz, A., White, B., White, S. and Yeh, T. VizWiz: Nearly real-time answers to visual question. In *Proceedings of UIST 2010*.
8. Brutlag, J. Speed matters for Google web search. June 22, 2009. http://bit.ly/IhsoTN
9. Büttcher, S., Clarke, C.L.A. and Soboroff, I. The TREC 2006 terabyte track. http://1.usa.gov/SWuEZD
10. Chen, J., Subramanian, L. and Toyama, K. Web search and browsing behavior under poor connectivity. In *Proceedings of CHI 2009*.
11. Chen, X. Bennett, P.N. and Collins-Thompson, K. and Horvitz, E. Pairwise ranking aggregation in a crowdsourced setting. In *Proceedings of WSDM 2013*.
12. Chilton, L.B. and Teevan, J. Addressing people's information needs directly in a web search result page. In *Proceedings of WWW 2011*.
13. Centintas, S. and Si, L. Exploration of the tradeoff between effectiveness and efficiency for results merging in federated search. In *Proceedings of SIGIR 2007*.
14. Crabtree, D.W., Andreae, P. and Gao, X. Exploiting under-represented query aspects for automatic query expansion. In *Proceedings of KDD 2007*.
15. Dean, J. and Barroso, L.A. The tail at scale. *CACM*, 56(2), 2013.
16. Dörk, M., Bennett, P. and Davies, R. Taking our sweet time to search. In *Proceedings of CHI 2013 Workshop on Changing Perspectives of Time in HCI*.
17. Dumais, S.T. Task-based search: A search engine perspective. Talk at *NSF Task-Based Information Search Systems Workshop*, March 14-15, 2013. http://bit.ly/15rK5tD
18. Faber, D. Google's Marissa Mayer: Speed wins. ZDNet. November 9, 2006.
19. Gan, Q. and Suel, T. Improved techniques for result caching in web search engines. In *Proceedings of WWW 2009*.
20. Garfield, E. Fast science vs. slow science, or slow and steady wins the race. *The Scientist*, September 17, 1990.
21. Google Webmaster Central Blog. Using site speed in web search ranking. April 9, 2010. http://bit.ly/acUf3Q
22. Hallnäs, L., and Redström, J. Slow technology – designing for reflection. *Personal Ubiquitous Computing, 5(3)*, 2001.
23. Hecht, B., Teevan, J., Morris, M.R. and Liebling, D.J. SearchBuddies: Bringing search engines into the conversation. In *Proceedings of ICWSM 2012*.
24. Hsieh, G. and Counts, S. mimir: A market-based real-time question and answer service. In *Proceedings of CHI 2009*.
25. Jeong, J.-W., Morris, M.R., Teevan, J. and Liebling, D.J. A crowd-powered socially embedded search engine. In *Proceedings of ICWSM 2013.*
26. Joachims, T. and Freitag, D. WebWatcher: a tour guide for the World Wide Web. In *Proceedings of IJCAI 1996*.
27. Kelly, D., Gyllstrom, K. and Bailey, E.W. A comparison of query and term suggestion features for interactive searching. In *Proceedings of SIGIR 2009*.
28. Kotov, A., Bennett, P., White, R.W., Dumais, S.T. and Teevan, J. Modeling and analysis of cross-session search tasks. In *Proceedings of SIGIR 2011*.
29. Lee, U., Kim, J., Yi, E., Sung, J. and Gerla, M. Analyzing answers in mobile pay-for-answer Q&A. In *Proceedings of CHI 2013*.
30. Lieberman, H. Fry, C. and Weitzman, L. Exploring the web with reconnaissance agents. *CACM, 44(8)*, 2002.
31. Morris, M.R., Teevan, J. and Panovich, K. What do people ask their social networks, and why? A survey study of status message QandA behavior. In *Proceedings of CHI 2010.*
32. Muhlke, C. Slow food. The New York Times, February 17, 2008.
33. Robertson, G. and Gao, X. Improving AbraQ: An automatic query expansion algorithm. In *Proceedings of Web Intelligence and Intelligent Agent Technology 2010*.
34. Shneiderman, B. Response time and display rate in human performance with computers. *Computing Surveys*, 16(3), 1984.
35. Shurman, E. and Brutlag, J. Performance related changes and their user impact. *Velocity 2009*. http://oreil.ly/fTmYwz
36. Smucker, M.D. and Clarke, C.L.A. Modeling user variance in time-biased gain. In *Proceedings of HCIR 2012.*
37. Smucker, M. D., and Clarke, C.L.A. Time-based calibration of effectiveness measures. In *Proceedings of SIGIR 2012*.
38. Tyler, S.K. and Teevan, J. Large scale query log analysis of re-finding. In *Proceedings of WSDM 2010*.
39. Thies, W., Prevost, J. Mahtab, T., Cuevas, G.T., Shakhshir, S., Artola, A., Vo, B.D., Litvak, Y., Chan, S., Henderson, S., Halsey, M., Levison, L. and Amarasinghe, S. Searching the World Wide Web in low-connectivity communities. In *Proceedings of WWW 2002*.
40. Wang, L., Metzler, D. and Lin, J. Ranking under temporal constraints. In *Proceedings of CIKM 2010*.
41. Wang, Y., Huang, X. and White, R.W. Characterizing and supporting cross-device search tasks. In *Proceedings of WSDM 2013*.
42. White, R.W. and Dumais, S.T. Characterizing and predicting search engine switching behavior. In *Proceedings of CIKM 2009*.
43. White, R.W., Richardson, M., Bilenko, M., and Heath, A.P. Enhancing web search by promoting multiple search engine use. In *Proceedings of SIGIR 2008*.
44. White, R.W. and Roth, R.A. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan Claypool, 2009.
45. Zhang, J., Ackerman, M. Adamic, L. and Man, K. QuME: A mechanism to support expertise finding in online help-seeking communities. In *Proceedings of UIST 2007.*