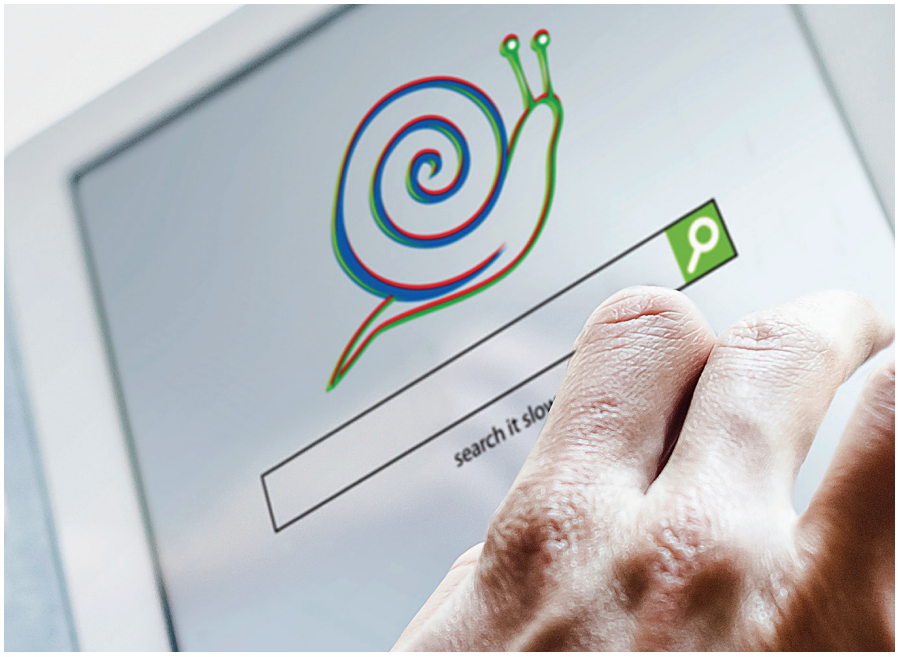
DOI:10.1145/2633041

Jaime Teevan, Kevyn Collins-Thompson, Ryen W. White, and Susan Dumais

Viewpoint
Slow Search

Seeking to enrich the search experience by allowing for extra time and alternate resources.

We live in a world where the pace of everything from communication to transportation is getting faster. In recent years a number of "slow movements" have emerged that advocate for reducing speed in exchange for increasing quality. These include the slow food movement, slow parenting, slow travel, and even slow science. Building on these movements we propose the concept of *slow search*, where search engines use additional time to provide a higher quality search experience than is possible given conventional time constraints.



Speed, Speed, Speed

Substantial research and engineering effort has been devoted to achieving low latency in large, complex computing systems such as search engines.[6] Search engines target speed for good reason. Research suggests that people perceive results that are delivered quickly as higher quality and more engaging than those delivered more slowly. Online experiments where server-side delays are injected into the delivery of search results have shown negative impact on people's search behavior.[11,12] For example, Google reported that intentionally increasing the load time of their search result page by as little as 100 milliseconds decreased the number of searches per person. Further, these differences increased over time and persisted even after the delays were removed. In similar experiments, Bing observed that artificial delays lead to a decrease in the number of queries and clicks, and an increase in time to click. Even improvements that seem like they should positively impact the searcher experience have been shown to have negative outcomes if they increase latency. For example, when Google experimented with returning 30 results instead of 10, they found that the number of searches and revenue dropped significantly because the additional results took a half-second longer to load.[10]

To achieve near-instantaneous speed, search engines make a number of compromises. They limit the complexity of the features and models used to identify relevant documents by, for example, making simplistic assumptions about language, often 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 aspects of the similarity between queries and documents. Time-saving mechanisms such as search-result caching and index tiering are also heavily exploited, despite the risk that such approaches may cause relevant content to be missed.

Not All Searches Need to Be Fast

Although searchers have grown accustomed to rapid responses to their queries, recent advances in our understanding of how people search suggest there are scenarios where a search engine could take significantly longer than a fraction of a second to return relevant content.[12] While someone search-

ing for a specific website or a straightforward fact almost certainly wants immediate results, people often invest minutes, hours, or even days in more complex or exploratory search tasks. A person planning a vacation or researching a medical diagnosis, for example, may be willing to wait for better results or insights. Additionally, since it is now possible to predict if an individual will resume a search task at a later date,[8] slow search tools could make use of the time between sessions to produce high-quality search results that could then be presented immediately when a search task is resumed.

Slow search approaches are valuable and often necessary when people have intermittent, slow, or expensive network connections. In such cases it can be difficult for searchers to employ traditional search strategies, such as rapidly reformulating queries. Instead, successful search systems must provide mechanisms that enable the systems to make the best use possible of the available time. For example, RuralCafe helps searchers in rural regions limit the number of iterations that they need to make by providing an expanded query interface and performing additional post-query processing which affects the type and richness of the response.[4] Mobile phones also often have limited bandwidth, and slower search processing times may be acceptable given that most of the delay a searcher observes is caused by network latencies in fetching data to the device.[9] Likewise, future space travelers may 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 unlikely the searcher would even notice the extra time invested.

Supporting Slow Search

Slow search techniques can be used to improve search quality over the course of seconds, minutes, or longer. With even just a little extra time to invest, search engines can relax existing restrictions to improve search result quality. For example, complex query processing can be done to identify key concepts in the query, and multiple queries derived from the initial query can be issued to broaden the set of candidate documents to cover different aspects of the query.[5] Search engines can also make use of additional time to employ resources that are inherently slow, such as other people. Crowd-based ranking methods use human judgments to identify the most relevant existing content for a query. For example, a Korean question-answering service Jisiklog allows mobile searchers to 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 quality of the responses.[9]

Slow search techniques can be used to improve search quality over the course of seconds, minutes, or longer.

While additional time can be used to identify particularly relevant results within the existing search engine framework, it can also be used to create new search artifacts and enable previously unimaginable user experiences. For example, researchers have used crowdsourcing to extract textual content from search results and synthesize that content into an inline answer for display on the result page.[3] Instead of merely returning a list of links, slow search results can include a summary or synthesis of the result content, the necessary background material to understand a topic, or the context necessary to resume an ongoing tasks. Rather than simply helping people find existing content, slow search systems can also facilitate the generation of new content that can be archived for use in future searches by the current searcher and others with similar interests.

Slow search systems require computational mechanisms to determine the appropriate method to employ given the time and resource constraints of a particular query. Recent research has examined tradeoffs between effectiveness and efficiency, and has developed

Calendar of Events

August 17–22
ACM SIGCOMM 2014 Conference,
Chicago, IL,
Sponsored: SIGCOMM,
Contact: Fabian E. Bustamante,
Email: fabianb@cs.northwestern.edu

August 21–22
Collaboration Across Boundaries: Culture, Distance & Technologies,
Kyoto, Japan,
Sponsored: SIGCHI,
Contact: Naomi Yamashita,
Email: yamashita.naomi@lab.ntt.co.jp

August 23–27
International Conference on Parallel Architectures and Compilation,
Edmonton, Canada,
Sponsored: SIGARCH,
Contact: Jose Nelson Amaral,
Email: jamaral@ualberta.ca

August 24–27
The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining,
New York, NY,
Sponsored: SIGKDD & SIGMOD,
Contact: Claudia Perlich,
Email: Claudia.perlich@gmail.com

August 24–27
International Conference on Parallel Architectures and Compilation,
Edmonton, Canada,
Sponsored: SIGARCH,
Contact: Jose Nelson Amaral,
Email: jamaral@ualberta.ca

September 1–4
25th ACM Conference on Hypertext and Social Media,
Santiago, Chile,
Sponsored: SIGWEB,
Contact: Leo Ferres,
Email: leo@inf.udec.cl

September 1–5
27th Symposium on Integrated Circuits and Systems Design,
Aracaju, Brazil,
Sponsored: SIGDA,
Contact: Edward David Moreno,
Email: edwdavid@gmail.com

computational models to support different policies.[2] For example, cascade models for time-sensitive ranking degrade effectiveness gracefully under time constraints.[13] Devoting a significant amount of additional resources to create a better search experience can be expensive, but slow search also presents cost-saving opportunities. Search engines currently experience significant demand on resources during periods of high activity, and must develop infrastructure to handle peak loads. If not all query processing needed to happen immediately, this load could be distributed more evenly to leverage underutilized resources.

Slow Search Implications

Slow search will change the way that people experience search, including how they express what they are looking for and how they interact with the information that they find. One source of inspiration for the slow search experience is question asking. People regularly wait for high-quality answers to the questions they ask online.[1] Although most search engines represent a searcher's need using a single short query, slow searchers may learn to express needs more richly. In online social situations people provide long natural language explanations of what they are looking for. 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?" Additional context can also be identified implicitly, as is often done in personalized search. Slow search engines need to clearly communicate the status of the slow searches to searchers and help them understand the benefit of a delayed system response. These systems should notify users when new relevant content is available, and provide ways for searchers to interrupt a slow search if it appears to be heading off-track or the results are no longer required.

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 on and learn about the topic of their search. Dörk et al.[7] 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 session. With additional time, search engines can help people comprehend the context of that information and learn what is necessary to fully understand it. A slower search process will not only allow search engines to identify and return the most relevant content, but may also enable searchers to get the most possible from the search experience.

Our hope is that slow search will inspire new and creative research into how the search experience can be enriched with a more nuanced notion of relevant search resources and time constraints. C

References

1. Aperjis, C., Huberman, B.A., and Wu, F. Human speed-accuracy tradeoffs in search. (Jan. 11, 2010); http://bit.ly/9bTehC.
2. Azari, D., Horvitz, E., Dumais, S., and Brill, E. Actions, answers, and uncertainty: A decision-making perspective on web-based question answering. *IP&M 40*, 5 (2004), 849–868.
3. 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*, (237–246).
4. Chen, J., Subramanian, L., and Li, J. RuralCafe: Web search in the rural developing world. In *Proceedings of WWW 2009*, 411–420.
5. Crabtree, D., Andreae, P., and Gao X. Exploiting underrepresented query aspects for automatic query expansion. In *Proceedings of KDD 2007*, 191–200.
6. Dean, J. and Barroso, L.A. The tail at scale. *Commun. ACM 56*, 2 (Feb. 2013), 74–80.
7. Dörk, M., Bennett, P. and Davies, R. Taking our sweet time to search. In *Proceedings of the CHI 2013 Workshop on Changing Perspectives of Time in HCI*.
8. Kotov, A., Bennett, P., White, R.W., Dumais, S.T., and eevan, J. Modeling and analysis of cross-session search tasks. In *Proceedings of SIGIR 2011*, 5–14.
9. 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*.
10. Linden, G. Marissa Mayer at Web 2.0. (Nov. 9, 2006); http://bit.ly/2J2amv.
11. Shurman, E. and Brutlag, J. Performance related changes and their searcher impact. Velocity 2009; http://oreil.ly/fTmYwz.
12. Teevan, J., Collins-Thompson, K., White, R.W., Dumais, S.T. and Kim, Y. Slow search: Information retrieval without time constraints. In *Proceedings of HCIR 2013*.
13. Wang, L., Lin, J., and Metzler, D. A cascade ranking model for efficient ranked retrieval. In *Proceedings of SIGIR 2011*, 105–114.

Jaime Teevan (teevan@microsoft.com) is a senior researcher at Microsoft Research and an affiliate professor at the University of Washington, Seattle, WA.

Kevyn Collins-Thompson (kevyn@umich.edu) is an associate professor at the University of Michigan, Ann Arbor, MI.

Ryen W. White (ryenw@microsoft.com) is a senior researcher at Microsoft Research, Redmond, WA.

Susan Dumais (sdumais@microsoft.com) is a distinguished scientist and deputy managing director at Microsoft Research, Redmond, WA.