ACM-W Athena Lecture — Large-Scale Behavioral Data: Potential and Pitfalls

Susan T. Dumais

Microsoft Research

Redmond, Washington, USA

Permission to make digital or hard copies of part or all 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 third-party components of this work must be honored. For all other uses, contact the Owner/Author. Copyright is held by the owner/author(s).

CHI'15 Extended Abstracts, Apr 18-23, 2015, Seoul, Republic of Korea
ACM 978-1-4503-3146-3/15/04.

<http://dx.doi.org/10.1145/2702613.2725426>

Abstract

Over the last decade, the rise of web services has made it possible to gather traces of human behavior in situ at a scale and fidelity previously unimaginable. Large-scale behavioral data enables researchers and practitioners to detect adverse drug reactions and interactions, to understand how information diffuses through social networks, how people browse and search for information, how individual learning strategies are related to educational outcome, etc. Using examples from search, I will highlight how observational logs provide a rich new lens onto the diversity of searchers, tasks, and interactivity that characterize information systems today, and how experimental logs have revolutionized the way in which web-based systems are designed and evaluated. Although logs provide a great deal of information about what people are doing, they provide little insight about why they are doing so or whether they are satisfied. Complementary methods from observations, laboratory studies and panels are necessary to provide a more complete understanding of and support for search which is increasingly a core fabric of people's everyday lives. The CHI community should lead the way in shaping best practices and policy in behavioral log studies.

Author Keywords

search; behavioral data; personalization; search strategy; user search behavior; query analysis; click analysis; search intent; methodology

ACM Classification Keywords

H.1.2 User/Machine Systems: Human Factors; H.3.3 Information Systems: Information Search and Retrieval.

Biography

Susan Dumais is a Distinguished Scientist at Microsoft and Deputy Managing Director of the Microsoft Research Lab in Redmond. Prior to joining Microsoft Research, she was at Bell Labs and Bellcore, where she worked on Latent Semantic Analysis, techniques for combining search and navigation, and organizational impacts of new technology. Her current research focuses on user modeling and personalization, context and search and temporal dynamics of information. She has worked closely with several Microsoft product groups (Bing, Windows Desktop Search, SharePoint, and Office Online Help) on search-related innovations. Susan has published widely in the fields of information science, human-computer interaction and cognitive science, and holds several patents on novel retrieval algorithms and interfaces. Susan is also an adjunct professor in the Information School at the University of Washington. She is Past-Chair of ACM's Special Interest Group in Information Retrieval (SIGIR), and serves on several editorial boards, technical program committees, and government panels. She was elected to the CHI Academy in 2005, an ACM Fellow in 2006, received the SIGIR Gerard Salton Award for Lifetime Achievement 2009, was elected to the National Academy of Engineering (NAE) in 2011, and received the ACM Athena Lecturer Award, and Tony Kent Strix Award in 2014.

Athena Lecturer Award

The ACM-W Athena Lecturer Award celebrates women researchers who have made fundamental contributions to Computer Science. Each year ACM will honor a preeminent woman computer scientist as the Athena Lecturer. Speakers are nominated by SIG officers. The Athena Lecturer will give a one-hour invited talk at an ACM conference determined by the speaker and the SIG which nominated her. A video of the talk will appear on the ACM website.

2014-15 Athena Lecturer: Susan T. Dumais

ACM-W named [Susan T. Dumais](#) of Microsoft Research as the 2014-2015 Athena Lecturer. Dumais introduced novel algorithms and interfaces for interactive retrieval that have made it easier for people to find, use and make sense of information. Her research, at the intersection of human-computer interaction and information retrieval, has broad applications for understanding and improving searching and browsing from the Internet to the desktop. The Athena Lecturer award celebrates women researchers who have made fundamental contributions to computer science. It includes a \$10,000 honorarium provided by Google Inc.

"Dumais has helped us understand that the search is not the end goal," said Mary Jane Irwin, who heads the ACM-W awards committee. "Her focus is on understanding when and why people search, and presenting results in context to help integrate those results into the larger search process. Her sustained contributions have shaped the thinking and direction of human-computer interaction and information retrieval, and influenced generations of student interns through collaborative projects with academic and industry partners."

Dumais' initial research demonstrated that different people use different vocabulary to describe the same thing, and that this mismatch limits the success of traditional keyword-based information retrieval methods. To build search systems that avoided the vocabulary problem, she and her colleagues invented Latent Semantic Indexing (LSI). A key feature of LSI is its ability to extract the latent conceptual structure from a large collection of texts by analyzing the associations between terms that occur in similar contexts, thus enabling a search engine to retrieve using concepts rather than keywords. Beyond information retrieval, LSI has been used to model various aspects of human cognition such as language acquisition and textual coherence.

Recently, Dumais' research has analyzed how web content changes over time and how people revisit web

pages, establishing that re-visitation patterns are influenced by user intent and changes in content. Her results have produced a retrieval model that uses web page changes to improve search ranking, and new tools to help people understand how the information they interact with changes over time in both expected and unexpected ways. Finally, her research on user modeling and context has enabled search engines to personalize search experiences for different individuals.

The author of more than 200 articles on information science, human-computer interaction, and cognitive science, Dumais holds several patents on novel retrieval algorithms and interfaces.

References

- [1] ACM-W Athena Lecturer Award.
<http://awards.acm.org/athena/> .