Interactions with Search Systems

Information seeking is a fundamental human activity. In the modern world it is frequently conducted through interactions with search systems. The retrieval and comprehension of information returned by these systems is a key part of decision making and action in a broad range of settings. Advances in data availability, coupled with new interaction paradigms and mobile and cloud computing capabilities, have created a diverse set of new opportunities for information access and use.

In this comprehensive book for professionals, researchers, and students involved in search system design and evaluation, search expert Ryen White discusses how search systems can capitalize on these new capabilities, and how next-generation search systems must support higher-order search activities such as task completion, learning, and decision making. He outlines the implications of these changes for the evolution of search evaluation, as well as related challenges that extend beyond search systems in areas such as privacy and societal benefit.

RYEN W. WHITE is a Principal Researcher at Microsoft Research. His research is focused on understanding search interaction and on developing tools to help people search more effectively. White's doctoral research received the British Computer Society's Distinguished Dissertation Award. He recently received the Karen Spärck Jones Award, for contributions to information retrieval. White's research is featured in many Microsoft products.

Interactions with Search Systems

Ryen W. White

Microsoft Research

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107034228

First published 2016

Printed in the United States of America by Sheridan Books, Inc.

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data
White, Ryen W., author.
Interactions with search systems / Ryen W. White, Microsoft Research.
pages cm
Includes bibliographical references and index.
ISBN 978-1-107-03422-8 (hardback)
1. Search engines – Technological innovations. I. Title.
TK5105.884.W53 2016
025.04252–dc23 2015026053

ISBN 978-1-107-03422-8 Hardback

Cambridge University Press
978-1-107-03422-8 - Interactions with Search Systems
Ryen W. White
Frontmatter
More information

To Henry, Hannah, and Resa.
Your love and patience made this possible.

Contents

Cambridge University Press
978-1-107-03422-8 - Interactions with Search Systems
Ryen W. White
Frontmatter
More information

Cambridge University Press
978-1-107-03422-8 - Interactions with Search Systems
Ryen W. White
Frontmatter
More information

Preface

Information seeking is a fundamental human activity, often conducted through interactions with automated search systems. The retrieval and comprehension of information returned by these systems is a key part of decision making and action in a broad range of settings; searching skills are now even taught in schools. The processes by which people retrieve and use information has been examined in detail by the information science, information retrieval, and human-computer interaction research communities for decades.

Information scientists have targeted the cognitive and behavioral mechanisms involved in the formulation of information needs and the processes by which people search for information and update their beliefs. The goal of searching is often regarded to be to reduce uncertainty in light of the information encountered, but there may also be the intention to increase that uncertainty, for example during exploratory or leisure search scenarios. Information retrieval researchers have targeted the development of new search technologies, including more advanced methods for ranking, indexing, and crawling, that facilitates the collection and selection of potentially relevant content from large document collections such as the World Wide Web or within large enterprises (where the goal may be to locate people with specific expertise rather than find information items). Human-computer interaction researchers have investigated how people interact with technology, and they have developed interfaces to allow searchers to explore and make sense of information resources as well as generate hypotheses to guide future exploration activities and decision making. In this book I discuss how new interaction capabilities such as touch and gesture, the emergence of cloud and mobile computing, machine learning, and big (and small) data mining will change the search landscape over the next decade and beyond. By building on these and other pillars, next-generation search systems will empower people and support the activities that they value.

This is the first book devoted to discussing how the range of emerging technologies can be employed to improve the search experience. To enable this transformation, many research communities – including information retrieval, human factors, data mining, and machine learning – must cooperate on the development of systems that empower searchers and leverage the broad array of tools at their disposal to make search a productive and pleasurable experience.

There have been three documented "revolutions" in information retrieval research: (1) *cognitive* (focusing on the intellectual processes involved in the search for information), (2) *relevance* (understanding the different types of relevance and the criteria associated with each type), and (3) *interactive* (the provision of search support and the capture and use of preferences from searchers, e.g., as relevance feedback). The interactive revolution continues to the present day. Together, we will cover many of the latest technological advances in this area as we progress through this book. Importantly, we are also in the early stages of a fourth revolution: the *data revolution*, driven by the increased size of corpora such as the Web and enhanced capabilities to record, analyze, and learn from aggregated usage data at scale across a broad range of searchers, tasks, and activities. Open data movements promote the availability of data, often non-textual material such as maps and genomes, for use, reuse, and redistribution. Governmental initiatives such as Data.gov (in the United States) and Data.gov.uk (in the United Kingdom) are making a broad range of datasets, tools, and applications in topics such as agriculture, health, and education freely available for download.

The data revolution also includes the collection and analysis of "small data" describing individuals in increasing detail across a range of platforms. Data in this context includes that stored in repositories collected by search providers (comprising information such as queries and resources accessed), the signals that are available to search systems through new interaction modalities such as touch and gesture, as well as the output of a range of signals available on modern devices, including those from physiological and motion sensors. These new sources of data enable search systems to better interpret the searcher and the search situation, helping them to better adapt to received requests and operate proactively on the searcher's behalf to identify and act on available information.

Popular commercial search engines such as Google and Bing have millions of users and serve search results for billions of search queries daily. Like never before, search providers have access to significant data about the search and browsing behavior of the general population and, through the sensor network created by logging these queries, to data about world at large. Richer sensing capabilities and the Internet of Things (Höller et al., 2014) also herald the availability of a range of signals from large quantities of noisy sensors. Combining data from all of these sources at massive scale has incredible potential for understanding the human condition and tracking populations, to benefit humankind through important applications in domains such as health care. Although logs lack annotations about aspects including search intent, success, experience, and attention,[1] in the aggregate they are useful for population-scale monitoring and predicting the future. With access to such data, scientists can study how people search and consume information resources, and make discoveries to improve people's lives and make the world safer. For example, log data has recently been applied to predict influenza (Ginsberg et al., 2009), detect evidence of adverse drug reactions / interactions between medications (White et al., 2013b), and predict disease outbreaks (Radinsky and Horvitz, 2013). As part of the contribution of this book, I discuss efforts to make these data and its derivatives more widely available in the interests of science, while still respecting the privacy of individuals whose data are gathered and mined (which should be of paramount concern).

The ongoing "big data" revolution in computer science (and social science and related disciplines) has been facilitated by advances in distributed computing algorithms and technologies. By employing sophisticated data mining methods, search engines can quickly identify relevance signals in logged search behavior and use these to improve the search experience by better ranking search results, suggesting related searches or query completions, or recommending content seemingly useful to others engaged in the same or similar search tasks. For example, the rate with which particular results are clicked can be used to determine dominant search intents for ambiguous queries, and for a given search query received many times by a search engine, the popularity of particular results can be used as a way to rank order them by estimated likelihood of meeting searchers' information requirements (Joachims, 2002). Behavioral data captured in this way can also be used to refine accepted models of interaction behavior and support the creation of user models and metrics to assess system performance in offline settings.

In addition to focusing on the queries and result selections on the search engine result page (SERP), the *trails* that people follow in document corpora such as the Web (which is becoming increasing pervasive) or other hypertext collections could be used in social navigation to help guide future searchers. These trails capture traces of human behavior in a way similar to that in the physical world (e.g., read and edit wear on documents, footprints in sand, and paths through landscapes). Leveraging these trails to support future searchers realizes the vision of Vannevar Bush in his seminal 1945 paper "As We May Think": the capture of our trails through online resources allows us all to be digital trailblazers (Bush, 1945). Applying templates to sequences of logged data enables search providers to harness the procedural search knowledge of the masses, allowing them to provide strategic search support spanning the full search task, rather than simply directing searchers to potentially useful resources.

Of course, considering that humans are affected by a range of cognitive biases, and search log data are mined from their online behavior, the data can be affected by biases in how people perceive the information that they encounter. These biases can skew the behavioral signals used to direct the individual (filter bubble effects) and future searchers via aggregated behavior (e.g., search behavior fueled by cognitive biases, behavioral biases, common misconceptions, and misinformation and rumor) that need to be handled. As a result, care must be taken in interpreting behavioral signals, and steps should be taken to de-bias them – for example, by uniformly sampling clicks to remove position effects, considering the role of caption content in driving click decisions, or employing mechanisms to remove known biases during the collection of data or the analysis of experiments (e.g., controlling for the impact of neighbor effects in controlled experiments on social networks [Eckles et al., 2015]).

Beyond storing, analyzing, and using interactions with existing technologies, new sources of data are also emerging as increasingly important. Cloud computing means that people's data are no longer segregated on different machines. Search providers often have multiple product offerings, with search being only one of these. Access to data from the other services that are typically offered (electronic mail, file sharing, productivity applications, etc.) can help improve search system effectiveness. This information could be available, with user consent, in personalized search applications, whereby search providers can model preferences based on contextual clues such as the documents they have reviewed or edited, as well as those from which searches are initiated. For example, a search for "VAR" from inside an Excel spreadsheet provides additional clues that the query relates to the Excel function for variance, and not about other interpretations of the acronym (e.g., value at risk, etc.). Contextual information is also available in Web search, with the queries and Web page visits that precede a search providing valuable additional relevance signals to better rank search results and generate content recommendations. Other signals such as location, time, and even the document collection itself (e.g., hyperlinks pointing to the document currently being read) also offer useful information about the current search situation. Recent advances in modeling spatial context could be useful in modeling interests and intentions within and between locations, or other applications such as supporting more effective human motion in the world (e.g., within-building travel directions).

Although I largely focus on Web search in this book, many of the lessons also apply in other domains that might be mediated through the Web but where the information is not part of the general Web. A significant amount of electronic searching is domain specific, including legal, medical, and intellectual property. These may use technology such as Web browsers as the platform, but the accessed material may not be in the form of Web pages. Even within the context of Web search itself, there are different verticals – images, video, news, and so forth – each of which has its own presentation formats and methods of interaction. For example, in image search, infinite scrolling through pages of image thumbnails is a common practice not observed in the examination of traditional search results comprising lists of Web pages. These specialized search scenarios may deserve their own search systems, but the boundaries are also blurred and content from search verticals bleeds into the presentation of traditional SERPs, affecting signals such as click-through rates on other SERP elements (e.g., advertisements [Metrikov et al., 2014]).

Social networking websites have emerged over the past decade and have grown rapidly. Data from these networks can be used to model people's interests and address issues of data sparseness for personalization, but also provide real-time support to searchers by connecting them with friends, acquaintances, or domain experts willing to assist them in a conversation in real time. Rather than conversing directly with others, the searchers could also leverage crowdworkers to help them with their information-seeking activities or other tasks. These workers can help find search results or compose instant answers that have immediate utility (Bernstein et al., 2012), but can also be stored in repositories for future searchers with similar interests.

There is also a growing interest in the development of anticipatory services in the form of intelligent agents such as Microsoft Cortana, Apple Siri, and Google Now – all of which look deeply into the searcher's context and can leverage the power of backend search systems to answer searcher questions. Cortana and Google Now can also proactively locate relevant and timely information based on people's preferences and other signals from implicitly and explicitly provided data. Over time, by anticipating searchers' needs, these systems will reduce the volume of queries that searchers must formulate and issue on their own accord.

Advances in data availability coupled with new interaction paradigms (touch, gaze, large displays, gesture, spoken dialog) and mobile computing capabilities (more powerful tablets, smartphones, and phablets [combining features of both device forms]) have created a broad range of new opportunities for information access and use. People can now interact with search systems in more lightweight and natural ways using modalities such as touch (swipes and pinches on phone and tablets), gesture (on devices such as the Kinect or the Leap), augmented reality, and more accurate speech recognition. All of these developments make interaction easier in situ, when people may be engaged in some other in-world task where more standard interaction methods such as keyboards and mice are not available. By being able to handle coarse-grained interactions such as swipes, scrolls, pinches, and other manipulations with a single or multiple fingers, devices can offer intuitive search interfaces that are robust to noise. Gestural interaction – coupled with the incorporation of large-screen displays – enables a compelling range of possible interactive scenarios in the immersive search arena. Recent advances in machine learning have enabled significant improvements in speech recognition technology to make it usable for unconstrained scenarios such as processing voice queries on mobile devices and real-time conversations with intelligent assistants. Background noise that can degrade speech recognition also provides useful contextual signals about the ambient environment that can be useful for ranking and recommendation applications.

Mobile devices such as smartphones and slates have become powerful and versatile. The integration of sensors such as accelerometers, gyroscopes, and proximity sensors provides rich context for applications. Indeed, evidence from self-reports and log analyses suggests that people can, and will, use search technology anytime, anywhere. The availability of these devices allows search systems to help people solve tasks in situations where previously there was no help available. Almost any question (or argument!) can now be resolved with a Web search on one's mobile device. Cross-device activities, where people transition between different devices (e.g., home computer to smartphone), occur frequently, and tasks can span multiple devices are now commonplace (Montañez et al., 2014). This creates new opportunities for systems to perform so-called *slow searches* to utilize the downtime between tasks or proactively support task resumption when the searcher does restart the search. For example, if a searcher terminates a task that they were attempting on their desktop computer, and the system can predict that they will resume it sometime later on their smartphone, it could use inter-task downtime to proactively retrieve content of possible interest, automatically or via third-party human involvement.

Moving beyond handheld devices such as smartphones, wearable and augmented reality applications allow relevant information to be incorporated into people's view of the world. Although this has been proposed in a range of scenarios, including triage in emergency medical situations and printer repair, recent advances in mobile computing, recommendation, and speech recognition offer further opportunities to enhance search interaction. One example are smart glasses (e.g., head-mounted displays such as *Google Glass*), which provide additional alerts on incoming communications, directions to destinations of interest, and location-dependent information, all in the searcher's field of view and using speech input. With such "always on" technology, systems no longer need to rely on their users to request information explicitly; relevant information can be pushed toward searchers proactively, capitalizing on signals such as their current location and historic preferences. Despite the attractiveness of such a seamless interaction method, there are privacy concerns around its use in public settings and the collection and capture of data (videos, still images, audio) about other individuals without their consent that need to be considered as part of broader discussions on the societal implications of wearable technology (Hong, 2013; Hoyle et al., 2014). Microsoft *HoloLens* is another example of augmented-reality technology housed inside a head-mounted display that supports multiple modes of interaction (including gaze and speech), and also adds three-dimensional holographic representations to the user's field of view, which can be manipulated using physical gestures.

Behind many of these technological advances lie conceptual models of human-information behavior, especially around how people seek and consume information. Although a number of influential models have been proposed over the past few decades, primarily by researchers in the information science community, these models must keep pace with technological advances and searcher demands. It is clear from the advances described so far that we are in the midst of a revolution in information seeking, in terms of both interaction capabilities and data availability. A plethora of new technologies and dedicated applications are emerging to supersede generic search systems in some information settings, and searcher demands for ubiquitous access to information continues to grow. In this book I discuss how these new capabilities will affect search system design, and how the range of evaluation methods and metrics that are applied to determine the performance of search systems needs to be extended.

In this book I focus on online searching. However, many of the concepts and ideas could also be applied in other domains, such as digital libraries or specialized domains. Regardless of the information environment, given its importance and technological advances, there is still significant opportunity for substantial impact through work on search interaction. Next-generation search systems will support a range of tasks, from finding basic facts to helping people explore, learn, and use encountered information. We are only beginning a journey to a more enlightened society facilitated by next-generation search technology.

The book is divided into four parts designed to introduce core concepts, describe models and methods, cover evaluation, and cast an eye to the future. The target audience for this book spans many readers, but it is primarily aimed at scholars at all career stages, including graduate students seeking to learn more about interactions with systems, especially how evidence is collected, modeled, and applied to improve the search experience for the current searcher and for future searchers. I also discuss the future of search interaction and highlight emerging trends that I believe will be of increasing importance in the design of search systems over the next decade and beyond.