

Reimagining Search

Search engine developers are moving beyond the problem of document analysis, toward the elusive goal of figuring out what people really want.

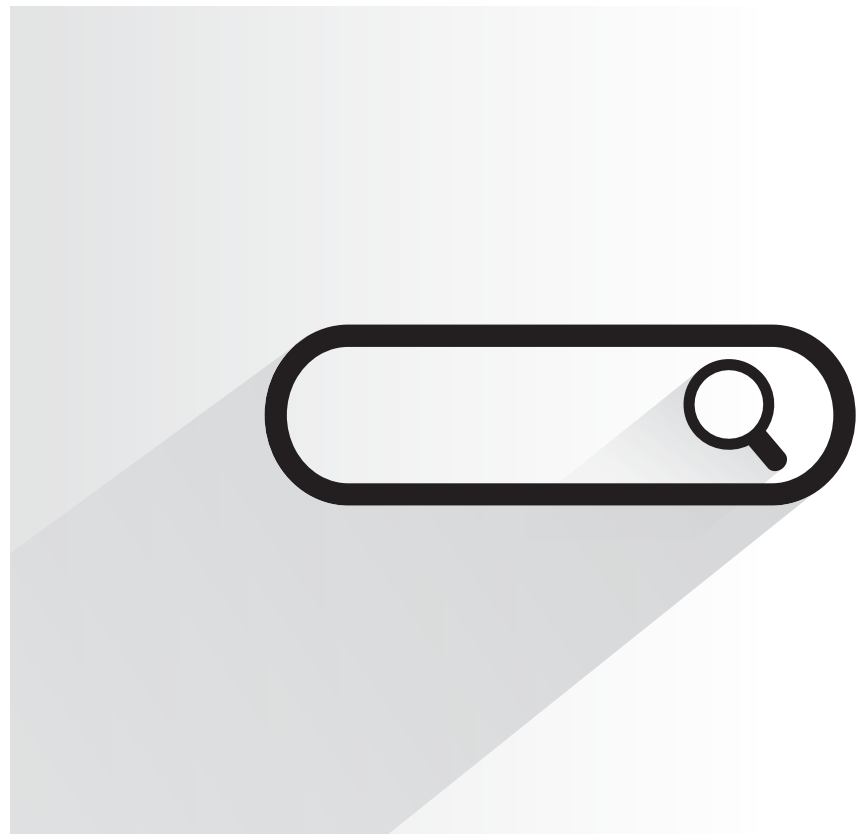
EVER SINCE GERARD SALTON of Cornell University developed the first computerized search engine (Salton's Magical Automatic Retriever of Text, or SMART) in the 1960s, search developers have spent decades essentially refining Salton's idea: take a query string, match it against a collection of documents, then calculate a set of relevant results and display them in a list. All of today's major Internet search engines—including Google, Amazon, and Bing—continue to follow Salton's basic blueprint.

Yet as the Web has evolved from a loose-knit collection of academic papers to an ever-expanding digital universe of apps, catalogs, videos, and cat GIFs, users' expectations of search results have shifted. Today, many of us have less interest in sifting through a collection of documents than in getting something *done*: booking a flight, finding a job, buying a house, making an investment, or any number of other highly focused tasks.

Meanwhile, the Web continues to expand at a dizzying pace. Last year, Google indexed roughly 60 trillion pages—up from a mere one trillion in 2008.

"As the Web got larger, it got harder to find the page you wanted," says Ben Gomes, a Google Fellow and vice president of the search giant's Core Search team, who has been working on search at Google for more than 15 years.

Today's Web may bear little resemblance to its early incarnation as a academic document-sharing tool, yet the basic format of search results has remained remarkably static over the years. That is starting to change, however, as search developers shift focus from document analysis to the even thornier challenge of trying to understand the kaleidoscope of human wants and needs that underlie billions of daily Web searches.



While document-centric search algorithms have largely focused on solving the problems of semantic analysis—identifying synonyms, spotting spelling errors, and adjusting for other linguistic vagaries—many developers are now shifting focus to the other side of the search transaction: the query itself.

By mining the vast trove of query terms that flow through Web search engines, developers are exploring new ways to model the context of inbound query strings, in hopes of improving the precision and relevance of search results.

"Before you look at the documents, you try to determine the intent," says Daniel Tunkelang, a software engineer who formerly led the search team at LinkedIn.

There, Tunkelang developed a sophisticated model for query understanding that involved segmenting incoming queries into groups by tagging relevant entities in each query, categorizing certain sequences of tags to identify the user's likely intent, and using synonym matching to further refine the range of likely intentions.

At LinkedIn, a search for "Obama" returns a link to the president's profile page, while a search for "president" returns a list of navigational shortcuts to various jobs, people, and groups containing that term. When the user selects one of those shortcuts, LinkedIn picks up a useful signal about that user's intent, which it can then use to return a highly targeted result set.

In a similar vein, a search for "Hemingway" on Amazon will return a

familiar-looking list of book titles, but a search for a broader term like “outdoors” will yield a more navigational page with links to assorted Amazon product categories. By categorizing the query—distinguishing a “known item” search from a more exploratory keyword search—Amazon tries to adapt its results based on a best guess at the user’s goal.

The widespread proliferation of structured data, coupled with advances in natural language processing and the rise of voice recognition-equipped mobile devices, has given developers a powerful set of signals for modeling intent, enabling them to deliver result formats that are highly customized around particular use cases, and to invite users into more conversational dialogues that can help fine-tune search results over time.

Web users can see a glimpse of where consumer search may be headed in the form of Google’s increasingly ubiquitous “snippets,” those highly visible modules that often appear at the top of results pages for queries on topics like sports scores, stock quotes, or song lyrics. Unlike previous incarnations of Google search results, snippets are trying to do more than just display a list of links; they are trying to answer the user’s question.

These kinds of domain-specific searches benefit from a kind of a priori knowledge of user intent. Netflix, for example, can reasonably infer most queries have something to do with mov-

By analyzing the interplay of query syntax and synonyms, Google looks for linguistic patterns that can help refine the search result.

ies or TV. Yet a general-purpose search engine like Google must work harder to gauge the intent of a few characters’ worth of text pointed at the entire Web.

Developers are now beginning to make strides in modeling the context of general Web searches, thanks to a number of converging technological trends: advances in natural language processing; the spread of location-aware, voice recognition-equipped mobile devices, and the rise of structured data that allows search engines to extract specific data elements that might once have remained locked inside a static Web page.

Consumer search engines also try to derive user intent by applying natural language processing techniques to inbound search terms.

For example, when a user enters the phrase “change a lightbulb,” the word “change” means “replace;” but if a user enters “change a monitor,” the term “change” means “adjust.”

By analyzing the interplay of query syntax and synonyms, Google looks for linguistic patterns that can help refine the search result. “We try to match the query language with the document language,” says Gomes. “The corpus of queries and the corpus of documents come together to give us a deeper understanding of the user’s intent.”

Beyond the challenges of data-driven query modeling, some search engine developers are finding inspiration by looking beyond their search logs and turning their gaze outward to deepen their understanding of real-life users “in the wild.”

“Qualitative research is great to generate insight and hypotheses,” says Tunkelang, who sees enormous potential in applying user experience (UX) research techniques to assess the extent to which users may trust a particular set of search results, or exploring why they may not choose to click on a particular link in the results list. Qualitative research can also shed light on deeper emotional needs that may be difficult to ascertain through data analysis alone.

At Google, the search team runs an ongoing project called the Daily Information Needs study, in which 1,000 volunteers in a particular region receive a ping on their smartphones up

Milestones

Computer Science Awards, Appointments

PAPADIMITRIOU AWARDED VON NEUMANN MEDAL

IEEE has honored Christos H. Papadimitriou, C. Lester Hogan Professor in the Department of Electrical Engineering and Computer Science at the University of California, Berkeley, with the 2016 John von Neumann Medal “for providing a deeper understanding of computational complexity and its implications for approximation algorithms, artificial intelligence, economics, database theory, and biology.”

Papadimitriou, who has taught at Harvard, the Massachusetts Institute of

Technology, the National Technical University of Athens, Stanford University, and the University of California at San Diego, is the author of the textbook *Computational Complexity*, which is widely used in the field of computational complexity theory. He also co-authored the textbook *Algorithms* with Sanjoy Dasgupta and Umesh Vazirani, and the graphic novel *Logicomix* with Apostolos Doxiadis.

The IEEE John von Neumann Medal is awarded for outstanding achievements in computer-related science and technology.

ACM CITES PERROT FOR VISION, LEADERSHIP

ACM has named Ron Perrot of the Queen’s University Belfast/Oxford e-Research Centre recipient of the 2015 ACM Distinguished Service Award “for providing vision and leadership in high-performance computing and e-science, championing new initiatives and advocating collaboration among interested groups at both national and international levels.”

Perrott was cited for providing vision and leadership in high-performance computing and e-science, championing new

initiatives, and advocating collaboration among interested groups at the national and international levels. He has been an effective advocate for high-performance and grid computing in Europe since the 1970s, working tirelessly and successfully with academic, governmental, and industrial groups to convince them of the importance of developing shared resources for high-performance computing at both national and regional levels.

Perrot is a Fellow of ACM, IEEE, and the British Computing Society.

to eight times per day to report on what kind of information they are looking for that day—not just on Google, but anywhere. Insights from this study have helped Google seed the ideas for new products such as Google Now.

Researchers at Microsoft recently conducted an ethnographic study that pointed toward five discrete modes of Web search behavior:

- **Respite:** taking a break in the day's routine with brief, frequent visits to a familiar set of Web sites;
- **Orienting:** frequent monitoring of heavily-used sites like email providers and financial services;
- **Opportunistic use:** leisurely visits to less-frequented sites for topics like recipes, odd jobs, and hobbies;
- **Purposeful use:** non-routine usage scenarios, usually involving time-limited problems like selling a piece of furniture, or finding a babysitter, and
- **Lean-back:** consuming passive entertainment like music or videos.

Each of these modes, the authors argue, calls for a distinct mode of on-screen interaction, “to support the construction of meaningful journeys that offer a sense of completion.”

As companies begin to move away from the one-size-fits-all model of list-style search results, they also are becoming more protective of the underlying insights that shape their presentation of search results.

“One irony is that as marketers have gotten more sophisticated, the amount of data that Google is sharing with its marketing partners has actually diminished,” says Andrew Frank, vice president of research at Gartner. “It used to be that if someone clicked on an organic link, you could see the search terms they used, but over the past couple of years, Google has started to suppress that data.”

Frank also points to Facebook as an example of a company that has turned query data into a marketing asset, by giving marketers the ability to optimize against certain actions without having to target against particular demographics or behaviors.

As search providers continue to try to differentiate themselves based on a deepening understanding of query intent, they will also likely focus on capturing more and more information about the context surrounding a partic-

ular search, such as location, language, and the history of recent search queries. Taken together, these cues will provide sufficient fodder for increasingly predictive search algorithms.

Tunkelang feels the most interesting unsolved technical problem in search involves so-called query performance prediction. “Search engines make dumb mistakes and seem blissfully unaware when they are doing so,” says Tunkelang.

“In contrast, we humans may not always be clever, but we’re much better at calibrating our confidence when it comes to communication. Search engines need to get better at query performance prediction—and better at providing user experiences that adapt to it.”

Looking even further ahead, Gomes envisions a day when search engines will get so sophisticated at modeling user intent that they will learn to anticipate users’ needs well ahead of time. For example, if the system detects you have a history of searching for Boston Red Sox scores, your mobile phone could greet you in the morning with last night’s box score.

Gomes thinks this line of inquiry may one day bring search engines to the cusp of technological clairvoyance. “How do we get the information to you before you’ve even asked a question?” ■

Further Reading

Bailey, P., White, R.W., Liu, H., and Kumaran, G., Mining Historic Query Trails to Label Long and Rare Search Engine Queries. *ACM Transactions on the Web*. Volume 4 Issue 4, Article 15 (September 2010), <http://dx.doi.org/10.1145/1841909.1841912>

Lindley, S., Meek, S., Sellen, A., and Harper, R., ‘It’s Simply Integral to What I do.’ Enquiries into how the Web is Weaved into Everyday Life, WWW 2012, <http://research.microsoft.com/en-us/people/asellen/wwwmodes.pdf>

Salton, G., *The SMART Retrieval System—Experiments in Automatic Document Processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, 2012

Vakkari, P., Exploratory Searching as Conceptual Exploration, Microsoft Research, <http://bit.ly/1N3rI3x>

Alex Wright is a writer and information architect based in Brooklyn, NY.

© 2016 ACM 0001-0782/16/06 \$15.00

ACM Member News

A “LITTLE DIFFERENT” CAREER TRAJECTORY



“It’s a little different,” says Julia Hirschberg, Percy K. and Vida L.W. Hudson

Professor of Computer Science and Chair of the Computer Science Department at Columbia University, of her career trajectory.

Hirschberg majored in history as an undergraduate, earning a Ph.D. in 16th century Mexican social history at the University of Michigan at Ann Arbor. While teaching history at Smith College, she discovered artificial intelligence techniques were useful in building social networks of 16th century colonists from “fuzzy” data. She soon decided computer science was even more exciting than history and went back to school, earning a doctorate in computer science from the University of Pennsylvania in 1985.

“None of my career decisions have been carefully planned. You often see opportunities you never dreamed would be possible.”

As a result of her thesis work, Hirschberg met researchers at Bell Laboratories. She went to work there in 1985, first working in test-to-speech synthesis, then launching the Human-Computer Interface Research Department in 1994, and moving with Bell to AT&T Laboratories.

Hirschberg started teaching at Columbia in 2002, and became chair of the Computer Science Department in 2012. Her major research area is computational linguistics; her current interests include deceptive speech and spoken dialogue systems.

“One of the things I think of when I tell young women about my career is that many opportunities arise,” Hirschberg says. “I never knew as an undergraduate that I would become a computer scientist, let alone chairing a computer science department at Columbia. You make some decisions, but they are not necessarily decisions for life.”

—John Delaney