New metasearch engine leaves Google, Yahoo crawling

March 25 2009

Weiyi Meng, a professor of computer science at Binghamton University, State University of New York, is hopeful that one day in the not-too-distant future, you'll be able to type a query into an online search engine and have it deliver not Web pages that may contain an answer, but just the answer itself. Credit: Jonathan Cohen

One day in the not-too-distant future, you'll be able to type a query into an online search engine and have it deliver not Web pages that may contain an answer, but just the answer itself, says Weiyi Meng, a professor of computer science at Binghamton University, State University of New York.

For instance, imagine typing in "Who starred in the film Casablanca?" The search engine would respond with "Humphrey Bogart and Ingrid Bergman."

Not impressed?

Try asking a more nuanced question, such as "What do Americans think of universal health care?" A search engine will create a report indicating trends in opinion based on what has been posted to the Web.

Search engines may eventually be used to conduct polling and even help sort fact from fiction, said Meng, who is helping to make such possibilities a reality, both through his research and as president of a company called Webscalers.

The way Meng sees it, big search engines such as Google and Yahoo are fundamentally flawed. The Web has two parts: the surface Web and the deep Web. The surface Web is made up of perhaps 60 billion pages. The deep Web, at some 900 billion pages, is about 15 times larger.

Google, which relies on a "crawler" to examine pages and catalog them for future searches, can search about 20 billion pages. Web crawlers follow links to reach pages and often miss content that isn't linked to any other page or is in some way "hidden."

Meng, along with researchers at the University of Illinois at Chicago and the University of Louisiana at Lafayette, has helped pioneer large-scale metasearch-engine technology that harnesses the power of small search engines to come up with results that are more accurate and more complete.

"Most of the pages on the deep Web aren't directly 'crawlable.' We want to connect to small search engines and reach the deep Web," he said. "That's the idea. Many people have the misconception that Google can search everything, and if it's not there it doesn't exist. But we should be able to retrieve many times more than what Google can search."

Not only can a metasearch engine probe deeper, it can also offer the latest information.

"In principle," Meng said, "small guys are much better able to maintain the freshness of their data. Google has a program to 'crawl' all over the world. Depending on when the crawler has last visited your server, there's a delay of days or weeks before a new page will show up in that search. We can get fresher results."

The concept is not new. In fact, the first metasearch engine was built in 1994.

"The big difference between our technology and the ones pursued by other people is that most of the other technologies do the metasearching on top of a small number of general-purpose search engines, such as Yahoo, Google or MSN," Meng explained. "We have a completely different perspective. We want to build large-scale metasearch engines on top of many small search engines."

The Web has millions of search engines at businesses, universities, newspapers and other organizations. Since 1997, and with continued funding from the National Science Foundation, Meng and his collaborators have found ways to run queries across multiple search engines and sort through the results.

Webscalers is based in the Start-Up Suite at Binghamton University's Innovative Technologies Complex, which is home to several young companies that have their roots in faculty inventions.

"If the Web keeps on growing, a company like Google may run out of resources to crawl all of those pages," said Vijay V. Raghavan, vice president of Webscalers and a faculty member at the University of Louisiana at Lafayette. "We won't have that problem. We will scale

much better."

Webscalers' technology could be useful for large organizations with many divisions. For example, Webscalers has developed a prototype that would allow a search of all 64 campuses in the State University of New York system as well as SUNY's central administration.

"People can use it to find collaborators," Meng said. "It could also help prospective students find programs they're interested in."

The technology could be adapted to large companies or even the government, Meng said.

Challenges for large-scale metasearch engines include determining which search engines are the best for a given query, automating the interaction with search engines as well as organizing the search results.

Meng hopes to build a grand metasearch engine one day that would integrate all of the 1 million small search engines into a single system. "There are still a lot of significant challenges in creating a system of such magnitude," he said, "but I am optimistic that such a metasearch engine can be built."

Try out the concept online

Webscalers has already launched several metasearch products:

The first is a news metasearch engine called AllinOneNews. Available at www.allinonenews.com , it connects to 1,800 news sources in 200 countries. That's the largest metasearch engine in the world.

Webscalers also offers MySearchView, a system that allows any user to create his or her own metasearch engine just by checking off a few

options at www.mysearchview.com .

Source: Binghamton University

Citation: New metasearch engine leaves Google, Yahoo crawling (2009, March 25) retrieved 25 April 2024 from https://phys.org/news/2009-03-metasearch-google-yahoo.html