

Search engines that learn from searchers

January 8 2010, By Bill Steele

(PhysOrg.com) -- New research aims to create search-engine software that can learn from users by noticing which links they click and how they reformulate their queries when the first results don't pay off.

You probably won't notice, but in the near future some search engines may start experimenting on you. It should be worthwhile, since the goal is to better understand what you were looking for and give you the best possible answers.

New research by Thorsten Joachims and Robert Kleinberg, associate and assistant professors of computer science, respectively, aims to create [search-engine](#) software that can learn from users by noticing which links they click on in a list of search responses, and how they reformulate their queries when the first results don't pay off.

The work is funded by a four-year, \$1 million grant from the National Science Foundation under federal stimulus funding, formally known as the American Recovery and Reinvestment Act (ARRA). The research will lead to methods that improve search quality without human guidance, especially on specialized Web sites such as scientific or legal collections or corporate intranets.

Joachims believes the work will have long-term benefits for the economy, invigorating the market for high-quality and focused search software. "I think there is a potential for commercial impact, improving quality and productivity," he said. In the short term, the project will fund at least two Ph.D. students for 4 years, and provide research positions for

undergraduate students.

As a demonstration, the researchers plan to create a new search engine for the physics arXiv Web site at Cornell, which contains thousands of papers in physics, mathematics and computer science, and possibly for other specialized collections.

"In several ways, providing search for small collections is more difficult than for the whole Internet. [Google](#), Yahoo! and Microsoft can spend a lot of manpower on engineering a good ranking function for the Internet. For small collections, this has to happen automatically via machine learning to be economical." Joachims explained.

Search is not a one-size-fits-all business: People searching specialized collections might use the same words in very different ways. Is "uncertainty," for example, about the location of subatomic particles, career choices, investment opportunities or romance?

"The key idea is have a search engine that gets better just by people using it," Joachims said. He and his collaborators have already created a search engine called Osmot -- the name is a play on "learning by osmosis" -- which draws on extensive research by computer scientists in machine learning. The problem the new research will address is that what the machine learns may be biased by the way it displays results.

Eye-tracking studies done in cooperation with Geri Gay, the Kenneth J. Bissett Professor and Chair of Communication, have shown that absence of a click on a result at, say, the 11th position on the list of returns may mean that the result did not fit the user's information need, but it may also mean that the user had given up scanning the list that far down. To get reliable feedback from clicks, the search engine needs to shuffle the order in which results are returned.

"There is a trade-off. On the one hand, you want to present the best ranking you know so far," Joachims explained. "On the other hand, the search engine has to do a bit of experimentation to be able to learn even better rankings in the long run. The key is to balance the tradeoff between presentation and experimentation in an optimal way."

This trade-off is similar to what a gambler faces in a casino and is called a "multi-armed bandit" problem. When playing a row of slot machines, each play gives you new information about how much that machine pays, but also costs you a quarter. The trick is to eliminate some machines when you're sure they won't pay off without spending more than necessary. Kleinberg's work on algorithms for solving such trade-off problems will be key to making search engines learn effectively.

Osmot is open-source software but still very much in beta. More information can be found at learnimplicit.joachims.org/.

Provided by Cornell University

Citation: Search engines that learn from searchers (2010, January 8) retrieved 2 May 2024 from <https://phys.org/news/2010-01-searchers.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--