

# Researcher develops personalized search engines

May 15 2012

---

(Phys.org) -- With little more than basic information about Web users' behavior – that is, the hyperlinks they click on daily and the content at those sites – Susan Gauch can build a better search engine. In information systems research, this work is known as “implicit” user profiling, meaning there are basic assumptions about user interest and intent based on the sites they frequent and the content they view.

Gauch, a professor of computer science and computer engineering at the University of Arkansas, has expertise in developing robust and personalized search engines, which she will contribute to the work of Hypothes.is, a project started by Dan Whaley, the coder and entrepreneur who built the first Web-based travel reservation system. Hypothes.is is trying to build a system of annotation for the Web. Based on a model of community peer-review, the system will be an open-source platform that will enable annotators to comment on individual sentences.

“Since the very beginning of the Web, there has been an issue of trust,” Gauch said, “because there has always been this ubiquitous ability for anyone to create and distribute information. What Hypothes.is is trying to do is build confidence and trust about information obtained on the Web. Yes, it is a form of peer review, but it won't be hierarchical or purely academic. Many details haven't been worked out yet, but the peer-review component will be determined by the annotator's reputation, which will be based on many demographic factors and will be constantly under review by other annotators.”

The Hypothes.is system will function as an overlay – similar to the “Track Changes” tool in Microsoft Word – on top of stable content, including news, blogs, scientific articles, books, terms of service, ballot initiatives, legislation and regulations, software code and more. As an overlay, the system will not require participation of the underlying site, and the content will reside on separate servers, Gauch said. She added that by enabling or disabling plug-ins, users will have the power to turn the service on and off.

The work of Hypothes.is will build on the unique expertise of people like Gauch. Her work in user profiling relies on queries and log-in applications, proxy servers and especially plug-ins and cookies. Plug-ins can be thought of as bonus features attached to primary or master software. They are extra components that add specific abilities. A computer cookie is a small text file with a unique identification tag placed on the user’s computer by a website. Among other purposes, these features allow researchers to track user behavior. Gauch’s task then is to take this information and build a personalized search.

“If we know something about you, then we can build a profile and use different functions to tailor searches specific to your interests,” Gauch said. “Technology allows us to see which sites users frequently visit, so we have a good idea what they’re looking for when they enter vague or ambiguous search terms.”

How does this affect the user? Gauch mentions various search terms that cause great confusion without at least some basic knowledge about the user’s interests. For example, consider the word “rock” as a search term. With no information about the user’s browsing history, a simple Google search finds sites related to gems, geology, music (bands, merchandise and the hall of fame and museum in Cleveland), wrestling, the actor (Dwayne Johnson), political action groups (Rock the Vote), county government (Rock County, Minnesota, and Rock County, Wisconsin)

and something called “Rock Your Phone.” The information about a user’s behavior allows Gauch to build profiles that so that musicians will be directed to sites that sell guitars, and geologists will find the U.S. Geological Survey site.

Provided by University of Arkansas

Citation: Researcher develops personalized search engines (2012, May 15) retrieved 25 April 2024 from <https://phys.org/news/2012-05-personalized.html>

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.