

Researchers classify Web searches

April 10 2008

Although millions of people use Web search engines, researchers show that – by using relatively simple methods – most queries submitted can be classified into one of three categories.

Jim Jansen, assistant professor in Penn State's College of Information Sciences and Technology, worked with IST undergraduate Danielle Booth and Amanda Spink, Queensland University of Technology, to find that Web search engine users are doing primarily informational, navigational or transactional searching.

Informational searching involves looking for a specific fact or topic, navigational searching seeks to locate a specific Web site and transactional searching looks for information related to buying a particular product or service.

The research was the first published work of its kind done using actual searching data, with the aim of real-time classification. Researchers analyzed more than 1.5 million queries from hundreds of thousands of search engines users. Findings showed that about 80 percent of queries are informational and about 10 percent each are for navigational and transactional purposes.

Jansen and his colleagues arrived at those results by selecting random samples of records and analyzing query length, the order of the query in the session and the search results. These fields helped the team develop an algorithm that classified the searches with a 74-percent accuracy rate.

"Other results have classified comparatively much smaller sets of queries, usually manually," Jansen said. "This research aimed to classify queries automatically.

"Our findings have broad implications for search engines and e-commerce if they can classify the user intent of queries in real time. This is why we wanted a computational undemanding algorithm," Jansen continued. "It proves the 80/20 rule that 80 percent of the cases can be achieved with these clear-cut methods."

The paper "Determining the informational, navigational and transactional intent of Web queries" will appear in the May 2008 issue of *Information Processing & Management*. The article is currently available online.

The Penn State researcher said he plans to continue this research using a more complex algorithm that will hopefully yield a 90-percent accuracy rate using similar searching criteria.

Source: Penn State

Citation: Researchers classify Web searches (2008, April 10) retrieved 25 April 2024 from <https://phys.org/news/2008-04-web.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.