Researchers predict click-through behavior in Web searches

March 11 2009

In the world of search engines, clicks mean cash, and in a sluggish economy, companies can benefit by maximizing click-throughs to their Web sites from search engines.

Jim Jansen, associate professor of information sciences and technology, Penn State, and colleagues, developed a measurement to gauge customer satisfaction with search engine results. They hope this metric will aid companies in optimization and marketing.

"In some sense, this study is a first step in using neural networks in the analysis of user-system interactions for Web search data," Jansen said. "This research explores the online behaviors of users so that commercial search engine companies can utilize the data to improve click-through rates by designing more efficient retrieval and ranking algorithms."

Jansen, along with Ying Zhang, industrial and manufacturing engineering, Penn State and Amanda Spink, Queensland University of Technology, used search logs from Dogpile.com to find the factors that predicted increased or decreased user click-throughs. They reported their results in the March issue of the Journal of the American Society for Information Science and Technology.

Based on neural network data, the researchers were able to identify nine factors that can help predict future click-through rates, with five having positive effects, four having negative effects and one having no effect. The positive factors were: number of records in a search, sum of listing

ranks, mean [query length](#), browser type and [query time](#). Negative factors were: number of organic links clicked, rate of vertical type, time of first query and log in time. User intent, whether they were looking for information, to make a transaction or to navigate somewhere, did not have a significant impact on predicting future click-throughs.

"From a practical point of view, the more that a user reformulates the initial query, the click-through will increase, although there may be individual queries where the user clicks on no links," Jansen said.

The researchers used neural networks because they are modeling tools that can capture relationships between input and output. They took results from a Dogpile search engine transaction log and crated input and output values for the neural network.

"Because click-through is based on each user, we grouped the records according to each unique IP address and cookie to determine a single user," Jansen said. "A huge data set was not necessary to train the neural networks because the principle of training is about how to use insufficient data to get necessary relationships between inputs and outputs."

The researchers found that more searchers clicked through early in the day and more searchers using Internet Explorer clicked through. They also found that searchers who clicked through more often, used longer than average queries, modified their queries more than the average searcher and searched for a longer period of time than average. They also noted that those searching on the Web clicked through more than those searching Audio, Video or Images.

Source: Pennsylvania State University ([news](#) : [web](#))