
Searching For A Better Match

By *Charles Babcock*, http://www.zdnet.com/intweek/
July 31, 2000 3:53 AM PT

Forrester Research conducted a simple search for "HTML" on IBM's Web site and got back 10 pages of listings - one for each time a programmer neglected to close the initial HyperText Markup Language bracket on a page of IBM content, leaving the incomplete "html" tag.

The results were useless to someone searching the IBM site for HTML-based tools and technologies, Forrester analyst Paul Hagen noted in his June report, Must Search Stink?

But they are typical of the things that can go wrong on even sophisticated Web sites as content builds up into hundreds of thousands of pages.

Likewise, a Forrester search of the single keyword "buy" on the Lucent Technologies' site yielded results that included the company's human resource documents that told employees how to buy extra vacation days. Any visitor finding this information could also follow the links embedded in the documents to pages that discussed Lucent's executive compensation policies, Hagen wrote in the report.

It's not too far-fetched to say that a surgeon's scalpel is needed to cut through the clutter of the Net. However, the tool offered the average Internet user, the keyword search engine, perhaps AltaVista, Lycos or Google, is just a dull knife.

But a new group of search tools, called matching engines, promises to more easily hook up users with the information they're trying to find on the Web. While most matching engines are under development, the products should be on the market soon and present the first real search option for many users.

As John Lehman, chief executive of Sageware, says, "Use alternatives to the keyword search engine? There is no other kind."

Sageware is the maker of the Sageware set of applications that sit on top of keyword search engines and generate libraries of results to make the keyword search more relevant. Sageware is one way of producing results closer to what users have in mind when they formulate their queries.

At the same time, search engine designers have concluded that the somewhat crude, but effective, recognition engine - or collaborative filtering engine - such as LikeMinds, now part of Macromedia, and Firefly, now part of Microsoft, may illustrate a new way to bring more precise results to the end user. The collaborative filtering engine matches a set of characteristics of the user to a group profile. Amazon.com, for example, uses a collaborative filtering engine to categorize a visitor to like-minded visitors, based on previous purchases and clickstream decisions. Based on the interests of that group, the recognition engine triggers its best estimate of which book or CD the visitor is going to be interested in next.

A new set of matching engine makers - such as Burning Glass Technologies and iXmatch.com - wants to apply this approach to the user's broad-based searches. Instead of matching a user to a complex set of characteristics of a group, they try to apply one set of complex characteristics to another in order to obtain a precise and relevant match.

"They're getting close to the Holy Grail of search engines, but they're not there yet," says Sean Luitjens, director of strategic

technology at Monster.com, the popular résumé and recruitment site, which is looking at the new tools.

It's easy enough for Monster to list jobs, collect r,sum,s and put out a career newsletter, but matching exactly the right person to the right job remains difficult. Monster relies on a standard keyword search engine, Microsoft's Index Server, to bring individuals and employers together, but Luitjens says he would like more precise capabilities.

"We want to create a better match than just keywords and locations," Luitjens says. For example, a director of technology is looking for a job as vice president of technology, but is ill-informed about what employers look for when they seek to fill that position. Consequently, the information he or she volunteered may not match the job sought, Luitjens says.

"We'd like to implement a matching engine on the basis of a more complex search or gap analysis," Luitjens says. A director of technology should be able to come to Monster, he says, "saying: 'I think I'm ready to move on, but can I be sure I'm really ready?' " A matching engine should be able to prompt him for level of experience and "soft skills," as well as the hard skills that he believes are most relevant, Luitjens says. The most appropriate match.

The typical Internet search engine, matching one keyword to another, can't give weight to different skills within a skills set or analyze the complex factors that go into the skills that a given company is looking for, says Ted Crooks, vice president of marketing at Burning Glass, maker of the Lens matching engine. Burning Glass was founded by three employees of HNC Software, the company that used artificial intelligence to generate the Falcon fraud detection system for credit-card companies. HNC has just invested $1 million in Burning Glass, Crooks says.

Lens is still under development; it is likely to be implemented at a major customer site by September and generally available late this year, Crooks says. No pricing has been set. Using Lens, a recruiting site is 4.5 times more likely to find the right candidate for the job based on its matching capabilities than it would be using a "vocabulary [keyword] search engine," he notes.

Burning Glass' objective, Crooks says, "is not to do a better job than a keyword search engine. Our goal is to do a better job than a human."

A matching engine can conduct a search based on many more factors than a collaborative filtering engine can, says George Karypis, computer science professor at the University of Minnesota in Minneapolis. The first implementations are likely to be on job sites, such as Monster or Guru.com, where employers are trying to match complex skills sets with submitted résumés. There is no reason, however, that matching engines couldn't be used for broader search purposes, such as finding the right home for a buyer or finding the right parts for an electronics design engineer.

IXmatch makes the iX Toolkit matching engine, which is also under development and looking for its first customer. IXmatch was founded by former employees of NetPerceptions, maker of the collaborative filtering engine used at Amazon.

Prakash Puram, president of iXmatch, says the heart of the matching engine isn't a text keyword search; rather, it resembles color coding that illustrates what small part of three large circles overlap each other. The matching engine uses "set theory, correlation of coefficients, regression analysis and data clustering" in the process of trying to match one set of complex characteristics with another.

Sageware's Lehman, whose product refines the traditional keyword search engine, is skeptical that the matching engine makers will implement superior search technology.

Job sites and other manipulators of complex sets of characteristics might as well use Sageware to build libraries of known results for specific, frequently searched topics. Sageware can index eXtensible Markup Language tags, so that the information in documents is more easily tracked, he says. Searches based on XML-tagged documents are more relevant and consistently precise, Lehman says, though only a small proportion of the information available has been structured with XML tags.

Forrester's Must Search Stink? report urged companies trying to provide good search capabilities on their sites to "adopt content management practices that support consistent tagging and XML structuring."

Crooks says he can't explain the effectiveness of Lens without giving away technology information that Burning Glass is trying to patent.