

IRRATIONAL INQUIRERS

Finding data with SEARCH ENGINES is frustrating and cumbersome.

What's next?

BY CHRISTOPHER LOCKE

NOT SO LONG ago, high blood pressure and using a search engine were not synonymous. But that was when things were simple—when most data was stored on individual PCs and searching it was relatively painless.

The Internet radically altered the search equation (and the stress level). And the once provincial boundaries of the PC now extend beyond spinning hard drives to millions of Web sites and an enormous trove of data. Trouble is trying to find the proverbial needle in the haystack with search technology that isn't up to the task.

Anyone who has used search engines knows they're stupid. They always seem to return a flood of irrelevant results. Moreover, they make it difficult for the user to narrow down the search without throwing out potentially important results.

With this continued inability to deliver the goods, search engines, once hailed as the savior of the desktop, are now viewed as barely adequate software that dumps a hideous mess onto your computer screen. Despite these shortcomings, search engines are fundamental to the future of desktop computing—they maintain order in a potentially chaotic world of information overload—so the demand for better and faster search engines will only get more intense.

So, how do search engines and tools fall short? First of all, search tools, like Ask Jeeves, aren't able to correct users misspellings. And they don't recognize synonyms like *laptop* and *notebook*, treating them as totally different things, thus leading to completely different results. Another problem is that search tools are generic and don't offer much in the way of customization.

This means most users are afforded the same search experience, which often leads to receiving responses to a query based more on popularity than on relevance. What might be relevant to one person, however, may be useless to someone else. "Why the hell do I care about relevancy percentage?" asks Paul Hagen, senior analyst at Forrester Research, an IT consultancy. "When a search tool says a hit has 78 percent relevancy, what does that matter to me if it isn't what I want?"

Though things seem pretty bleak, all is not lost. There are several startups and public companies offering new software solutions to overcome the assorted problems. And some companies are even going further and using Extensible Markup Language (XML) to improve search technology. XML is able to index and place tags, or short descriptors, next to textual information so that machines can better determine the nature of that text.

One of the biggest complaints about search engines is that they don't understand simple requests. Go to just about any search engine and type in "Basketballs for less than fifteen dollars." *Red Herring* did. The results ranged from yesterday's NFL scores to an article on travel in Mexico to this strange one: "NCIA: Basic Criminal Justice Facts." Whatever happened to the basketball?

That experience isn't uncommon. According to the Boston Consulting Group, a market research firm, only 1.8 percent of all visitors to online retail sites actually make a purchase. In part, that's because nearly half the potential customers are unable to find the product,

which leads to an estimated loss of \$20 billion in sales. (For more on Internet searches and e-commerce behavior, see “A Questioning Look,” page 102.)

The good news is that several companies are developing software that would make search engines better able to understand queries. One of those companies, EasyAsk, a privately held company in Littleton, Massachusetts, uses the concept of natural language: a customer can type in questions in plain English and search across a variety of nonuniform sources.

ARCHIE'S BUNKER

Of course, search problems aren't new. There have been improvements along the way. One notable milestone in search engine development occurred in 1990 with the creation of Archie by Alan Emtage, a student at McGill University in Montreal who wanted to find a better way to search. Before Archie, the primary method of downloading files onto a desktop computer's hard drive was with file transfer protocol (FTP), a standard protocol that is part of the TCP/IP suite of Internet protocols. Client systems and servers used FTP to exchange files across TCP/IP networks in much the same way as peer-to-peer computing. Archie combed FTP sites and indexed all the files it found, achieving what at the time was the best and most popular repository of Internet files.

Unfortunately, not much has improved over the last 11 years. One recent improvement, however, is the advent of the modal-based search, which “learns” from prior searches and can thus better pinpoint the type of result a user wants. This goes beyond existing types by trying to understand the search visitor's needs. It was developed at the Xerox Palo Alto Research Center over two years ago. Among the companies using modal-based searches is Outride, which says its technology will find what you want twice as fast as a nonmodal search engine. The only drawback is that users need to fill out user and content profiles—and then the two have to get acquainted. One of Outride's competitors is Purple-Yogi. But unlike Outride, PurpleYogi offers free downloadable software that works in conjunction with a browser. What's clever about its software is that it continually updates a user's profile based on the sites and data the user seeks out on the Web.

TAGADELIC, BABY!

Many experts consider XML important for improving the world of search engines. XML, like HTML, makes use of tags, but while HTML specifies what each tag means, XML uses tags to add context to the data. Developed in 1996 under the auspices of the World Wide Web Consortium, XML is a set of rules for designing text formats for structured data, like spreadsheets, address books, and technical drawings. HTML is able to handle this information. But it can't transfer its structure; XML can. That's why XML is so powerful. But there are hitches—updating databases with XML is costly.

XYZFind, a venture-backed startup in Seattle, is fully aware of these shortcomings, but believes it's only a matter of time before XML is widely adopted. The company's software stores and updates any well-formed XML without user intervention or preparation. Moreover~ XYZFind returns actual content aggregated from all matching documents, not just links to documents.

Though XML is being used in news-service searches, some feel this is the wrong market to exploit its benefits. Factiva, a Dow Jones & Reuters company providing global news and business information, employs XML and is second only to Lexis-Nexis in its sector. Factiva says its services focus on intelligent indexing, and users will receive only highly relevant and structured data. But the fact that less than 3 percent of time-sensitive material is available in XML is considered unacceptable to Factiva's rival, Moreover.com, which doesn't rely on XML structured data, so is able to deliver result much faster than Factiva. “Since Factiva relies on licensed material,” says Nick Denton, CEO and founder of Moreover.com, “that takes time to index.”

Search engines are evolving quickly as corporations and end users push startups and incumbent vendors to improve the technology. What remains to be seen is if any of these companies can follow through on their promises an help restore some order to an increasingly chaotic world.