

## Bits and Bytes

### SEARCH ENGINES NOT ALL THEY'RE CRACKED UP TO BE

Michael Devitt

Using a search engine to surf the World Wide Web is becoming more and more like looking at an iceberg - you can see the tip, but you don't see everything that's just below the surface. According to a study published in the July 8th issue of *Nature* magazine, the most comprehensive search engine covers only about one-sixth of the total number of web pages that search engines can reach, and the top 11 search engines combined reference only 42 percent of the pages available on the Net.

Drs. Steve Lawrence and C. Lee Giles at the NEC Research Institute in Princeton, New Jersey gauged the Web's content and the popularity of pages by randomly sampling 2,500 IP addresses. They then visited those sites and surveyed and categorized them manually.

Giles and Lee estimated that the Web consists of about 800 million pages containing some 15 terabytes of data. Most of that data remains untapped by the Web's major search engines, leaving millions of pages of potentially important information virtually invisible to users who rely on those search engines to find what they're looking for.

Table I: Internet search engines and the percentage of the World Wide Web that they cover.

*Northern Light: 16.0% (out of 100)*

*Snap: 15.5%*

*AltaVista: 15.5%*

*HotBot: 11.3%*

*Microsoft: 8.5%*

*Infoseek: 8.0%*

*Google: 7.8%*

*Yahoo: 7.4%*

*Excite: 5.6%*

*Lycos: 2.5%*

*Euroseek: 2.2%*

Source: Lawrence S, Giles C. Accessibility of information on the web. *Nature* July 8, 1999.

The most comprehensive search engine, according to the researchers, is Northern Light ([www.northernlight.com](http://www.northernlight.com)), which indexes only 16 percent of the total pages on the Web. Other popular search engines such as Excite, Lycos and Yahoo each index less than half that amount.

The study also found that it takes an average of more than six months for a new web page to make it onto a search engine's listings. For some e-commerce sites, delayed indexing or not being listed on a search engine could make the difference between success and failure.

"That could have a substantial impact on their economic viability," Lawrence said. "The situation now is relatively unequal, in the sense that ... the more well known sites are the ones getting indexed."

Delayed or biased indexing could also influence a user's personal decisions, the researchers note.

"As the Web is being used more and more for different things, like people looking up political candidates for example, then if the index is biased, perhaps it is going to have an impact on the decisions people make," Lawrence said.

---

### Life in the Slow Lane

If you're a professional cowboy, trying to ride a 2,000 pound bull for the required eight seconds can seem like an eternity. According to a report from Zona Research, for some people, waiting for a web page to download can seem just as long.

The report, "The Need for Speed," reviewed the preferences of consumers and researched the economic impact of "bailout rates" for users with slower internet connections. Zona estimates that consumers with slow modem connections who are unable to load web pages within eight seconds often get frustrated, "bail out" and turn their attention elsewhere. E-commerce sites could face as much as \$73 million in lost revenue every month because of web page loading failures, while the securities and travel industries could lose up to \$40 and \$34 million annually due to unacceptable download times.

If you have a website that is graphic-intensive, or an online store that features large images of your products for sale, you may want to consider sacrificing a few graphics in favor of quicker download times. Your site may look incredible, but nobody's going to wait around if it takes 15 minutes to download.

Ask your patients if they are on line and what type of connection they have (dialup access, cable modem, T1 line, etc.). If you have not yet created a website but are thinking of doing so, gathering information in advance will help decide how your site will look and how patient-friendly it will be.

---

### A Net Divided

Although more Americans are getting on line every day, the gap between certain groups of people who have access to the Net and those who don't appears to be widening. A report from the National Technical Information Association predictably found that people living in rural areas and those making less money do not have the same access to the Net as upper-income households or people living in urban areas.

The report, *Falling through the Net: Defining the Digital Divide*, surveyed 48,000 households regarding points and methods of internet access, types of internet use, and reasons for discontinuing use, among other topics. Estimates show that although 40 percent of American households were on line last year, there is a growing trend in which whites are on the Net in a much greater percentage than African-Americans, Latinos or those of Asian descent.

The report's executive summary states: "For many groups, the digital divide has widened as the information 'haves' outpace the 'have nots' in gaining access to electronic resources. Until every home can afford access to those resources, however, we need public policies and private initiatives to expand affordable access to those resources."

The gap between the number of white versus non-white users and upper-income versus lower-

income users has grown in the past five years. African-American and Latino households are only about 40% as likely as white households to have Net access, and one-third as likely as those of Asian-Pacific Islander households. Whites are more likely to have internet access from home than blacks or Latinos have from any location (home, work, school, etc.).

Money is another factor in internet access. Households with incomes over \$75,000 per year are more than 20 times as likely to have access to the internet as those at the lowest income levels, and are more than nine times as likely to have a computer at home. And the divide between high and low-income users with internet access widened by 29 percent between 1997 and 1998.

To remedy the situation, the report suggests the continued use and growth of community access centers (CACs), such as libraries, schools and other access points. Providing public access to the internet, they believe, will help groups advance economically and provide them with the technical knowledge to compete in today's changing society.

"Establishing and supporting community access centers, among other steps, will help ensure that all Americans can access new technologies," the report states. "As we enter the Information Age, access to computers and the internet is becoming increasingly vital. It is in everyone's interest to ensure that no American is left behind."

Copies of *Falling through the Net* are available for free from the NTIA's website ([www.ntia.doc.gov/ntiahome/digitaldivide](http://www.ntia.doc.gov/ntiahome/digitaldivide)). Users can also download earlier editions of *Falling through the Net* and read related reports from the Department of Commerce.

AUGUST 1999