BY CAROL TENOPIR

Is Google the Competition?

LIBRARIANS, PUBLISHERS, and aggregators alike often call Google their main competitor. Google, or similar web search engines, is the information finding tool of first choice for many users—far ahead of proprietary online services or libraries and light years ahead of print sources.

The 2004 annual meeting of NFAIS (National Federation of Abstracting and Information Services) asked who would win the "battle for mindshare" in the year 2010. Will libraries, A&I (abstracting and indexing) services, and traditional publishers still exist, or will Google become the only information resource?

The first choice

Studies such as the Pew Internet and American Life study show that a majority of students first turn to search engines for school assignments. They recognize that not all information found there is reliable, but there is enough to satisfy most of their needs. What may be more surprising is that subject experts also often favor the web.

John Regazzi, managing director of market development at Elsevier, described studies that show 70 percent of professionals use the Internet in their work (even 91 percent of those over age 55 use it six or seven times each week). When Elsevier researchers asked librarians and scientists to name the top three most reliable online services, librarians named ScienceDirect, ISI's Web of Science, and Medline. Scientists, on the other hand, named Google, Yahoo!, and PubMed.

Making sense of data

Speakers at NFAIS offered a variety of tactics to deal with the current reality and preferences. Regazzi believes the winners will know what researchers need and can deliver it. Researchers are under pressure to move from research to product development faster. They require a wider range of sources. To meet these needs, researchers need data integration, not just Search. Internal and external sources must be integrated, scientific and business information needs to be searchable from the same Search, and data-mining tools should help identify trends and issues.

> One expert suggests leasing the Google search engine with "Google inside"

Some institutions are developing these new tools. Jane Rosov, coordinator, Data Distribution Program, National Library of Medicine (NLM), reported an increased interest in leasing the NLM databases for text mining. From only 25 customers (mostly aggregators) in 1985, NLM now leases Medline to 219 customers. These customers develop tools that reveal relationships among drugs, genes, and diseases; interfaces to allow better searching; and text mining to identify emerging diseases or biothreats.

If you can't beat 'em

Allowing search engines to index proprietary information is another tactic. Jan Pedersen, chief scientist at Yahoo!, described how search engines get content. The first is to crawl billions of free web pages. The second is to acquire feeds directly from content providers or permission to crawl fee-based services. Information stalwarts such as ProQuest, OCLC, and Reed Business Information now have some of their information indexed by Google, although access to full texts still typically requires a license. Nearly 100 percent of the Institute of Physics (IoP) content is indexed by Google as well as 90 percent of Emerald Insight publications. The American Institute of Physics (AIP) took a different route and built a new web site that is web crawler friendly.

In some disciplines, a Google search will now retrieve a fair amount of scholarly articles. Simon Inger, director, Scholarly Information Strategies in the UK, compared searching for known-item journal articles on Google and major A&I services. Physics articles were readily found on Google. In addition to AIP and IoP materials, the well-known physics e-print server arXiv.org has been indexed by Google for two years.

But business and social science articles are not widely indexed by Google. When they are, they are not typically within the first few result screens because social science search words are more common. Inger reported that CrossRef is looking to work with Google to index all of its content. "Expect all primary literature to be indexed in Google after that," he said.

Do what Google doesn't

Ben Shneiderman, professor at the University of Maryland and author of *Leonardo's Laptop* (MIT Pr., 2002), recommended several strategies for dealing with Google. Instead of (or in addition to) partnering with Google to index content, lease the Google search engine for proprietary systems with "Google inside." This will be restricted to what the publisher or library chooses to provide.

Finally, "do what Google doesn't do." Include multilingual and translation services that are better than automatic translations. Incorporate visualization techniques into search output. Extend linking and display into categorization (not just dynamically generated categories like those in the Vivisimo search engine but proven categories used in standard classification schemes). Provide multimedia products that mix types of information. And, speaking to both publishers and librarians, "focus on your specific communities of users" and "build on those communities" with special services and interactivity just for them.

Carol Tenopir
(ctenopir@utk.edu)
is Professor at the School of Information Sciences, University of Tennessee, Knoxville