The Tower of Google I

Supposedly, when the builders tried to reach heaven by building the Tower of Babel, they were prevented from doing so because they began to speak different languages. Their pride was punished not by raining fire and brimstone down on the monument, but by simply disrupting communications. Now there is an initiative underway to enhance the ability to communicate between seekers of information. This time, the Internet search company Google is attempting to construct another structure, albeit an electronic one, similar to another monument of the ancient world, the Library of Alexandria. What might be called the Tower of Google is an effort to digitize all of the information in the world and put it on the Internet in a searchable format. The effort is composed of two projects: the Print Publisher Program and the Print Library Project.

On his web site (http://www.policybandwidth.com/doc/googleprint.pdf) Jonathan Band, an intellectual property lawyer, describes these two programs:

Under the Publisher Program, a publisher controlling the rights in a book can authorize Google to scan the full text of the book into Google's search database. In response to a user query, the user receives bibliographic information concerning the book as well as a link to relevant text. By clicking on the link, the user can see the full page containing the search term, as well as a few pages before and after that page. Links would enable the user to purchase the book from booksellers or the publisher directly, or visit the publisher's website. Additionally, the publisher would share in contextual advertising revenue if the publisher has agreed for ads to be shown on their book pages. Publishers can remove their books from the Publisher Program at any time. The Print Publisher Program raises no copyright issues because it is conducted pursuant to an agreement between Google and the copyright holder.

Under the Print Library Project, Google plans to scan into its search database materials from the libraries of Harvard, Stanford, and Oxford Universities, the University of Michigan, and the New York Public Library. In response to search queries, users will be able to browse the full text of public domain materials, but only a few sentences of text around the search term in books still covered by copyright. This is a critical fact that bears repeating: for books still under copyright users will be able to see only a few sentences on either side of the search term. Users will *not* see a few pages, as under the Publisher Program, *nor* the full text, as for public domain works. Indeed, a full page of the book is never seen for an in-copyright book scanned as part of the Library Project unless a publisher decides to transfer their book into their Publisher Program account, in which case it would be under the agreement between Google and the copyright holder.

The objective is so far-reaching and captivating that is it hard to see how anyone could not support it: online access to content of the world's libraries. As consumers of knowledge, the possibility that we can find any paper or any book that we need for our work is tremendous. Even more so, there is the possibility of finding articles that we did not know existed. It could prevent us from either duplicating earlier work or correcting work that we found to be flawed in some way. In effect, the intellectual layer of the Earth, which Pierre de Chardin called the noosphere, becomes a lot denser.

More practically, it could resurrect older works that have gone out of print or have been relegated to publishers' backlists. As I noted recently ("Keeping Old Texts Alive," May 2005 editorial) a laser text that I co-authored in 1978 contains chapters that are out of date. But there were other chapters that are still relevant and would be useful at the undergraduate level. If the text were searchable and available in chapter-size sections, its useful life might be extended.

When I started as a young researcher here at Georgia Tech, I would go over to the Tech library each week and leaf through those journals to which I did not subscribe. I would return with a fistful of index cards and a number of Xerox copies of the most interesting papers. Now, it is possible to do a fairly complete search using the Web, but it still takes an effort to troll the journal sites to find the interesting papers, and then hop over to the eJournals section of the Tech library Web site to look at their contents. If this new Google Print project is realized, my search of a fully indexed set of journals each week for new journal content, based on my interests and keywords, would yield far more new ideas and results. How could anybody be against such an enterprise? But publishers and authors are up in arms over the prospect. In March, I'll discuss the possible consequences of this Tower of Google on users, publishers, authors, and societies like SPIE.

Donald C. O'Shea
Editor