OCLC Project Opens WorldCat Records to Google
by Barbara Quint

October 27, 2003 — OCLC has announced to its library members that it will begin testing the opening of WorldCat records to Google access. The project will extract a 2 million record subset from the more than 53 million records in the WorldCat database. The subset will target the most popular and widely available books by only selecting records with a minimum of 100 libraries holding each item. Searches on Google will retrieve the records and link through OCLC to library holdings. The move expands the scope of the Open WorldCat yearlong pilot project to make library resources available from non-library Web sites and will "test the effectiveness of Web search engines in guiding users to library-owned materials." (Details for potential participants can be found at http://www.oclc.org/worldcat/pilot/default.htm.) The Google/OCLC connection should be active by mid- to late November.

Over 12,000 academic, public, and school libraries that have contributed holding records or cataloging information to OCLC will automatically participate in the program. Corporate library collections will not be assumed to participate. Libraries may withdraw from the project by notifying regional service providers. Libraries that are not already part of the pilot project that want to participate can do so by joining OCLC or an OCLC cooperative and indicate their desire to join the pilot. Until the launch of the Open WorldCat pilot, the only end users that could reach WorldCat records were required to use FirstSearch, most probably in a library setting.

The announcement urged OCLC member libraries to participate in the program in order to increase the visibility of library collections to current and potential patrons, as well as enhance the image of libraries to administrators and funding agencies and improve the quality of material accessible from the Web. OCLC began testing Web outlets in September 2001 when it opened access to WorldCat records through links to selected online bookseller sites (e.g. BookPage.com, ABE Books, Alibris for Libraries, and the Antiquarian Booksellers’ Association of America). Of the four named, only BookPage offers a "find it at a library" for any and all searches; the rest only show the library link when the search has failed to find the item in the bookseller’s inventory. Currently, these sites generate some 50,000 library clickthroughs a month.

After users locate an OCLC record citation from a Google search or other Open WorldCat Web sites, clicking on the citation will link to an interface that requests a ZIP or postal code, state, or province. This in turn will locate the nearest libraries holding the item. Country names will suffice for locations outside the U.S. and Canada. The interface will also link to local library Web sites for further information, such as connections to the local library’s OPAC catalog or a map with driving directions. OCLC supplies all the links for libraries and at no extra charge. Once in a local library’s Web site, users may search for other items in individual library collections, but there is no option to search all of WorldCat.

OCLC launched the project in June 2003 after a year of extensive consultation with library and information industry leaders. This consultation bore out the perception that libraries needed more visibility on the Web, in particular to
reach people who don’t use their services now. Analysis of the pilot project will involve extensive feedback and surveys from participants and focus groups, as well as looking at click-through statistics. Participating libraries will receive usage statistics for the pilot project as part of OCLC’s analysis of its success. In June 2004, OCLC will decide whether to expand, continue, or discontinue the pilot project.

Clearly, the expansion to include OCLC’s records fits Google’s mission statement to “organize the world’s information and make it universally accessible and useful.” A Google spokesperson pointed to this as part of the company’s effort to reach beyond the open Web, as it has, for example, in including IEEE abstracts. Reports circulate that Google is in negotiation with several major international publishers as well.

There’s one problem, however. The programming algorithms that rank search results involve linguistic frequency as well as popularity statistics. OCLC bibliographic records, by Google spidering standards, are very thin. At this point, Google had nothing to say on how it will handle the OCLC records to ensure a “page one” level of visibility to searchers that corresponds to the quality of the material.

For the record, such issues may not take first place in Google’s in-house relevancy rankings for a while. Google executives are interviewing representatives of several investment banks preparatory to launching a projected $15 to $25 billion IPO next year. A story in the Financial Times said that Google might set up an online auction of shares so that millions of users could bid to own their own piece of the world’s most popular search engine. Danny Sullivan of SearchEngineWatch pointed to a Reuters report that only 53 public offerings have ever exceeded $1 billion, though Amazon and Yahoo! currently have a value of between $22 and $26 billion.

Barbara Quint, co-editor with Paula J. Hane for NewsBreaks, is editor in chief of Searcher, a columnist for Information Today, and a longtime online searcher. Her e-mail address is bquint@mindspring.com.