Key Words: From Publishing Afterthought to Invaluable Database

Speaker and Chair:
Iain E P Taylor
National Research Council of Canada
Research Journals
University of British Columbia
Vancouver, British Columbia

Speakers:
Gretchen P Purcell
Duke University Medical Center
Durham, North Carolina

Craig Van Dyck
John Wiley & Sons Inc
New York, New York

Reporter:
Mary Carr
USDA Forest Service
Olympia, Washington

Iain E P Taylor introduced the session on key words by comparing an index with a restaurant menu, in which not only the words but also the organization, content, and visual clues are important for finding what you want.

"Whether you call it an 'index' or a 'database', you still are referring to a 'menu'. To find out what's there, you need to know what to look for. The information needs to be presented in such a way that you can find it" said Taylor.

A usable index is based on the reader's expectations, not necessarily on the author's intentions, said Taylor. The reader comes to an article with a vocabulary and purpose different from the author's, so key words for the reader might not be the same as those for the author. Taylor suggested that the title, abstract, and full text all should be consulted to select appropriate key words that will let the reader know both what is there and what it's about.

With that, Taylor set the stage for a discussion of two new electronic indexing systems designed to make information more accessible to readers.

Putting Things in Context

Gretchen P Purcell, surgery resident and adjunct assistant professor of medical informatics at Duke University Medical Center and consultant for the *Annals of Internal Medicine*, suggested that the use of key words for indexing is inadequate to provide what the reader wants, because key words alone cannot provide the context in which the information resides. When the terms are in the wrong context, Purcell noted, they are more likely to lead to irrelevant retrievals.

Purcell presented a context-based markup language for indexing clinical articles in medical journals. This system captures the semantic context of terms by annotating the full text with both the key words and the types of information contained in the article. An electronic indexing system can then tailor a reader's search precisely. This approach, but not the same model, could be used for other types of articles such as case reports.

Purcell noted that uniform annotation can be done quickly by nonspecialists and result in consistency in indexing. Her technique was developed for clinical articles, but the model could also apply broadly to other scientific fields, especially research and case-study articles. See *GOTOBUTTON BM_1_ www.annals.org* for more information.

Connecting the Dots

Craig Van Dyck, vice president for production and manufacturing at John Wiley &

Sons, discussed CrossRef, which is designed to help readers find an article through a metadata database that cross-links online scholarly journals. The system is available to primary publishers that are members of the nonprofit Publishers' International Linking Association and to other users.

Participating publishers assign a unique and permanent identifier, the digital object identifier (DOI), to each article and deposit a set of bibliographic metadata into the database. Publishers tag the reference lists in their articles with SGML and then look up the DOIs of cited articles in the database, thereby linking their users to articles that are in the system—expected to exceed 3 million items this year. At a minimum, the cited publisher will provide bibliographic information; most provide abstracts, and some offer the article itself. Van Dyck emphasized that, unlike Purcell's context-based system, CrossRef retrieves only the references themselves; it is not designed to look up topics ("resource discovery").

Free to the end user, CrossRef enables multiple linking among publishers through one protocol and adds value to journals by helping readers go directly to cited references. See *GOTOBUTTON BM_2_www.crossref.org* for more information.

What's on the Menu?

The speakers agreed that indexing—whether based on key words or more—is all about retrieval of information to meet a reader's needs. "The key", said Taylor, "is to know what the reader wants. What should be on the menu?"