

Enhancing online bibliographic records to improve retrieval of reference collection monographs*

By Ruth H. Makinen, M.L.S.
Head of Technical Services

Betsy Friesen, M.L.S.
Library Specialist

University of Minnesota
Bio-Medical Library
Minneapolis, Minnesota 55455

Health sciences librarians frequently need to locate specific current information when providing reference service to clients. Over the years, reference librarians have devised various methods to help them quickly locate the correct title when needed. In the past, a separate card file, arranged by subject, containing records for reference monographs, helped the staff at the University of Minnesota Bio-Medical Library locate reference titles. The change to an online catalog provided new opportunities, enabling the staff to explore means to improve access to the reference collection and to also dispense with the onerous routine of producing and filing cards for that collection.

While considering ways to increase access to the reference collection, correspondence on the Public Access Computer Systems Listserv called attention to the merits of increasing subject retrieval by enhancing bibliographic records with content information. The authors decided to undertake a project to enhance the approximately 1,000 titles in the reference collection at the University of Minnesota Bio-Medical Library.

LITERATURE SEARCH

In his 1989 paper on access to the catalog record in the age of automation, Duke discusses the need for a revised cataloging code to take advantage of searching methods such as free-text searching with Boolean operators, truncated word matching, and proximity and adjacency searching. In his discussion of the catalog of the future, Duke describes a three-tiered record structure: (1) the document surrogate, the bibliographic citation that represents characteristics of the document and is similar to today's standard bibliographic record; (2) the document guide, a synopsis of the contents; and (3) the document text, the complete text online [1].

* An expanded version of a paper presented at the Ninety-Second Annual Meeting of the Medical Library Association, Washington, D.C., May 20, 1992.

The storage capacity of many of today's automated systems could become depleted quickly if Duke's approach were adopted in full. However, several bibliographic record enhancement projects have been reported that contain elements of Duke's catalog of the future, particularly the document surrogate and document guide tiers. Atherton's Subject Access Project of 1978 was the first proving ground for enhanced bibliographic records. She augmented standard bibliographic records with subject descriptors taken directly from books' contents and indexes to improve retrieval rates through online catalog free-text searching. The augmented records enhanced users' ability to locate subject information in the online catalog [2].

In 1987, Markey and Calhoun reported on the use of tables of contents and summary notes with unique words to enhance subject access in online catalogs. The addition of these fields increased the potential retrieval rates for natural-language keyword searching [3]. In 1988, Byrne and Micco found that enriching bibliographic records with content terms was "a viable and cost-effective technique for dramatically increasing the number of subject access points to the contents of books without a serious increase in false-drops" [4].

The literature of the 1990s contains many reports of experiments in enhancing standard bibliographic records with content information [5-9]. Overall, reviews have been mixed regarding bibliographic record enhancement. Dillon and Wenzel warn that although the addition of content information improves the overall retrieval, most of the improvement is in recall (the percentage of relevant documents in a collection retrieved for a query) at the expense of a decline in precision (the percentage of the retrieved documents that are relevant) [10]. Van Orden is also cautious and recommends that librarians "avoid unnecessary extravagances that will reduce efficiency" [11].

Finally, the Committee on the Machine-Readable Form of Bibliographic Information (MARBI) has issued two discussion papers proposing structural changes to the U.S. machine-readable cataloging record to better accommodate content data within the record [12-13].

PROJECT METHODOLOGY

With the insight gained from the literature review, the authors embarked upon a project to enhance online reference collection records with meaningful tables of contents or summaries. These fields are indexed for keyword searching in the University of Minnesota Libraries' online catalog.

In working through the reference collection for the retrospective portion of the project, librarians printed

out the bibliographic record for each title and used it as a worksheet. The value of the work in meeting reference needs was determined, and the bibliographic record was studied in light of these strengths and needs. If the librarian determined that all or part of the table of contents could be used to improve retrievability of the record, the table was photocopied, and the sections to be included in the enhanced record were highlighted for data entry. A summary also was created and added to the worksheet as necessary. The record-enhancement worksheets then were forwarded to a data-entry operator for input. A cataloging assistant reviewed the enhanced online record for errors. The same procedure was followed for new titles added to the collection.

Occasionally, an unhelpful main table of contents was followed by a very helpful list of appendixes or tables. Only information considered useful for the purpose of reference access was included in the record. Information on the end papers also was examined. When detailed tables of contents were too long to fit into one bibliographic record (record length could not exceed 7,000 characters, or bytes), only partial contents were included. Individual chapter authors are included in the contents note as space permits, particularly if an author is from the University of Minnesota.

Tables of contents frequently contained information that was introductory in nature or redundant. In these cases, content information was provided in a summary. These average 180 characters (four lines) and contain unique words, often taken directly from the volume, that describe the work and provide additional access. The summaries often are taken from the work's introduction or preface. The purpose of the summary is to improve the retrieval of the record via keyword searching, not just to provide a helpful description of the work.

As the project proceeded, the team found that 30% of the titles in the reference collection, although monographs, were cataloged in the serials format, with a note in the online catalog reading "latest edition in Bio-Med reference." Because the contents-note field is not used in the serials format, content information was included in the summary.

Several experiments were attempted with available equipment to scan table-of-contents data into the bibliographic record to avoid some of the manual data entry. The effort was abandoned, because the online catalog would not accommodate data scanned directly into the record.

Due to uncertainties regarding implementation of a "location-based searching" component within the catalog, a local note field containing the term *biomref* was added to each record. This enabled keyword searches to be restricted to the reference collection through addition of the phrase *and biomref* to each

search statement (e.g., "K = <KEYWORD> AND BIOMREF").

RESULTS

The retrospective portion of the project enhanced 1,100 records. The total will change as new titles are added and older volumes removed from the reference collection. Of the 1,100 records, 33% were enhanced with tables of contents, 39% with summaries only, and 20% with both a table of contents and summary. Eight percent of the records were not enhanced, because access would not have been improved.

The professional staff spent an average of seven minutes on each enhancement, for a total of 128 hours for the project. Photocopying and highlighting tables-of-contents data moved along rapidly. The most time was spent creating summaries. The data-entry operator averaged ten minutes per title, for a total of 183 hours. A library assistant also spent twenty-eight hours reviewing the online records for errors. Adding these figures, the estimated cost of the record enhancement project was \$4,342.00, or approximately \$4.00 per title.

Enhancement of new reference titles (approximately five titles per month) will continue as part of the regular cataloging effort. Retrospective enhanced records were contributed to the Research Libraries Information Network (RLIN); new enhanced records were contributed to both RLIN and the OCLC Online Computer Library Center.

Some extra maintenance is required for the enhanced bibliographic records. The records for new volumes or editions that are treated serially are evaluated to determine if they cover the same material as the previous edition or volume. The summaries of these records are edited as changes in coverage occur. When a title is removed from the reference collection, the record enhancements are not deleted; only the "biomref" note is deleted. An original record enhancement occasionally is augmented if requested by reference staff.

Reference staff use of the enhanced records was surveyed for two one-week (five-day) periods during the academic quarter. An average of 2.0 successful keyword searches restricted to the reference collection were conducted per day during the first week's survey, and an average of 2.2 per day during the second survey period. Several unsuccessful attempts also were recorded.

DISCUSSION AND CONCLUSION

The addition of the content information increases keyword retrieval rate. Recall is managed by limiting the search to information within the reference collection. As forewarned by Dillon and Wenzel, the

staff is finding that precision declines to some extent due to the addition of the content-bearing data. For example, a search "K=AIDS AND BIOMREF" yields sixty-four references. Although most of the titles recalled contain data about the disease, fifteen of the records are retrieved because they contain the word *aids* in a context such as "diagnostic aids" and thus are irrelevant. Overall, for this project, the benefits of adding the content-bearing data to the records outweighs problems with precision.

In addition to the retrieval benefits, including chapter-level data in the bibliographic record allows catalog users to determine if the work is suitable for their purposes, without going to the reference shelf. A streamlined form of enhanced cataloging may be added to other portions of the collection because of these advantages.

The use of keyword searching to locate items within the reference collection appears to be increasing. If the library continues to provide "ready reference" assistance to clients, and, as the reference staff becomes more familiar with the keyword searching capabilities offered by the addition of the content-bearing data, this additional form of access could become an integral part of reference work. The reference staff is employing the technique enthusiastically and is very positive about its capabilities for assisting clients.

It is important to remember that keyword searching does not and should not replace subject searching. The content-bearing data does not provide the collocation offered by a controlled subject vocabulary. The purpose of this project was to increase retrieval options by providing additional access points, not to replace the subject access already incorporated into the standard bibliographic record.

Is the addition of the content-bearing information cost-effective? The staff has been able to absorb the ongoing work into their routine, because the scope of this project is small, and no additional authority or analytic work is necessary. But the emphasis on enriching bibliographic records specifically for reference desk activities clearly increases the time spent on each record. This in-depth attention to the bibliographic record may not be feasible for a larger collection. A streamlined record enhancement project, in which meaningful content data were scanned directly into the bibliographic record according to established guidelines, similar to the Byrne and Micco experiment [14], could prove economical and achievable for a larger collection.

ACKNOWLEDGMENTS

The authors wish to thank Casidy Jones for her data-entry work, Jon Easley for his assistance with record

scanning, and Kathy Robbins for her enthusiasm and support of this project.

REFERENCES

1. DUKE J. Access and automation: the catalog record in the age of automation. In: Svenonius, E. The conceptual foundations of descriptive cataloging. San Diego: Academic Press, 1989:117-28.
2. ATHERTON P. Books are for use: final report on the subject access project to the Council on Library Resources. Syracuse, NY: Syracuse University, 1978.
3. MARKEY K, CALHOUN K. Unique words contributed by MARC records with summary and/or contents notes. In: ASIS '87: Proceedings of the 50th ASIS Annual Meeting. Medford, NJ: Learned Information, 1987:153-62.
4. BYRNE A, MICCO M. Improving OPAC subject access: the AFDA experiment. *Coll Res Libr* 1988 Sep;49(5):432-41.
5. DILLON M, WENZEL P. Retrieval effectiveness of enhanced bibliographic records. *Libr Hi Tech* 1990;31(3):43-6.
6. VAN ORDEN R. Content-enriched access to electronic information: summaries of selected research. *Libr Hi Tech* 1990;31(3):27-32.
7. MICHALAK T. An experiment in enhancing catalog records at Carnegie Mellon University. *Libr Hi Tech* 1990;31(3):33-41.
8. KNUTSON G. Subject enhancement: report on an experiment. *Coll Res Libr* 1991;52(1):65-79.
9. WEINTRAUB TS, SHIMOGUCHI W. Catalog record contents enhancement. *Libr Res Tech Serv* 1993;37(2):167-80.
10. DILLON, op. cit.
11. VAN ORDEN, op. cit.
12. MARBI. Content-enriched and enhanced subject access in USMARC records. Washington, DC: MARBI, 1990. (MARBI Discussion Paper no. 42).
13. MARBI. Enhancing USMARC records with table of contents. Washington, DC: MARBI, 1991. (MARBI Discussion Paper no. 46).
14. BYRNE, op. cit.

Received July 1992; accepted August 1994