Information Technology and Libraries, 8(1) 1989 *pp. 79-82.*

http://www.ala.org/index.cfm

Automated Control of Subject Headings at the OSU Libraries

Sally A. Rogers

Sally A. Rogers is Catalog Planning Librarian, Ohio State University Libraries.

As more libraries acquire online catalogs, the interest and attention given to automated authority control continues to increase. Baer and Johnson recently did a survey of the authority control literature produced since 1974. They found that most of it (some seventy references) did not support the idea that advanced computer-searching capabilities are making authority control less necessary in online catalogs. In addition, their survey of the uses of authority control in American college and university libraries inspired "a number of comments to the effect that maintaining authority control online was just as time-consuming as maintaining authority control manually."[1] If automation has not made authority control unnecessary or even less time-consuming, what *has* it done in this area? Baer and Johnson express the hope that automation has enabled libraries to provide *better* authority control even if it has not allowed them to save time.[2] Such has been the case at the Ohio State University Libraries (OSUL). What follows is a description of how automation has recently been used to improve authority control of subject headings in the online catalog.

Since December 1981, the OSUL online catalog (LCS) has included an authority file, which also serves as an index to the bibliographic records in the catalog[3] and consists of four headings indexes that provide access by author, subject, series, and uniform title. The Library of Congress name and subject authority tapes have been used to add headings and cross-references to the LCS authority file and to update the data for consistency with current LC practice.[4] (Markey and Vizine-Goetz's research report *Characteristics of Subject Authority Records in the Machine-Readable Library of Congress Subject Headings*[5] should prove invaluable to any library planning to use the LC subject tape.)

A tape containing OCLC records created or edited by OSUL is added to LCS each week by batch processing. At that time, headings on the records being added to the catalog are machine-matched against headings already in LCS—headings tagged 600, 610, 650, etc. are matched against the LCS subject index; those tagged 100, 700, etc. are matched against the name index; and so forth. Certain types of headings are reported for human review. If errors are found, LCS is corrected manually.

Regular review of these reported headings indicated that some kinds of errors were repeatedly occurring. Since few of the headings on OCLC records used for copy cataloging are checked before the records are added to LCS, the same outdated headings and subdivisions kept reappearing in the subject index. If an incoming heading exactly matched a cross-reference in LCS, it was automatically "flipped" to the correct form; but there was no provision for automatically correcting only *part* of a heading when it was added to the catalog.

OSUL staff responsible for heading maintenance kept track of the errors that occurred most frequently and wrote specifications for a computer program to correct as many of them as possible automatically. Several lists of outdated geographic names and topical subdivisions were

compiled and incorporated into the program. Since November 1987 this headings control program (HCTL) has been applied weekly to headings being added to LCS. In March 1988 it was retrospectively applied to all headings in the LCS subject index.

The HCTL program represents an extension of the automatic control made possible by the use of the LC authority tapes. It provides the capability to delete or replace heading subdivisions as specified on three lists that are part of the program. "List A" includes subdivisions to be deleted from all headings (e.g., —Addresses, essays, lectures). "List B" includes more than thirty geographic names to be replaced by newer forms, whether they appear as main headings or subdivisions (e.g., Zaire replaces Congo). "List C" includes more than sixty abbreviated or outdated subdivisions to be replaced (e.g., —Pol. & govt, becomes —Politics and government and —Yearbooks becomes —Periodicals). These lists can be amended as additional problems are identified and as LC policy changes.

HCTL makes it possible for part of a heading to flip to another form while any subdivisions following it remain the same. For example, "Insurance, Social—Florida" can be partially flipped to "Social security—Florida" if the cross-reference "Insurance, Social, see Social security" is already in LCS. If the cross-reference points to more than one heading, HCTL reports the possible partial flips so that library staff can decide which one is most appropriate.

Even when automatic correction is impossible, HCTL assists OSUL staff by identifying headings likely to be incorrect and reporting them for human review. A fourth list ("List X") was used in the retrospective application of HCTL to identify headings with "city flip" subdivisions (i.e., subdivisions no longer authorized for use after names of cities). These headings were reported for review and manual correction.

The program also isolates problems by looking for a partial match when an incoming heading does not exactly match any heading already used in LCS. It truncates the heading from the right end by dropping the final subdivision (e.g., Agriculture— Research—United States becomes Agriculture—Research). The search for an exact match is then repeated using the truncated form, or "grandparent" heading. If no exact match is found, the heading is further truncated by dropping a second subdivision from the right end (e.g., Agriculture—Research becomes Agriculture) and so forth, until no subdivisions remain. The headings for which there are still no exact matches found at the end of the truncation process are said to have "No grandparent in LCS" (i.e., the unsubdivided heading does not match any other LCS heading). The headings that wind up in this category are the ones most likely to contain errors, particularly typographical ones (e. g., Agriculture).

HCTL further isolates headings likely to contain errors by looking at their verification status during the truncation process. Headings in the LCS indexes are coded as either verified or unverified; those that are unverified are flagged with an asterisk. Headings become verified by meeting certain criteria; for example, any heading on an OCLC record created and input by LC according to *AACR2* is automatically verified when it is added to the file. Headings already in LCS become verified when they match those on the LC authority tapes or when they are manually verified by OSUL staff.

Separate reports are generated for headings that have verified and unverified grandparents. For example, if no exact match is found in LCS for the incoming heading "Dogs—Breeding," the HCTL program truncates the heading and looks for "Dogs." If an exact match is found and "Dogs" is verified on LCS, then "Dogs—Breeding" is said to have a verified grandparent in LCS. Headings with unverified grandparents are more likely to contain errors than those with verified ones.

In a limited number of cases, headings with verified grandparents become verified themselves without being reported for review. OSUL staff identified 14 very general subdivisions (e.g.,—Periodicals) that may be added to virtually any heading. If a heading consists of a verified grandparent plus one of these subdivisions, it is automatically verified without being reported.

While the HCTL program was written primarily to control subject headings, it has also been applied to personal name headings to identify and, in some cases automatically combine, two headings that are identical except for the lack of a death date in one. The program looks at the verification status of both headings and corrects one of them or simply reports them, according to specifications.

When the HCTL program was retrospectively applied, there were approximately 766,000 headings (including cross-references) in the subject index. These headings came from cataloging done by OSUL on OCLC since 1972 and from catalog records of the State Library of Ohio, Center for Research Libraries, and the U. S. Government Printing Office that had been added to LCS. Table 1 shows how many of these headings were automatically corrected and/or reported for review.

These figures show that 6 percent of the headings in the LCS subject index in March 1988 were automatically corrected by the HCTL program. An additional 7 percent were reported as potential problems needing review and manual correction. The status of 18 percent of the headings on the file changed from unverified to verified because the grandparent headings were verified.

On a weekly basis an average of 1,250 OCLC records are added to LCS, and the HCTL program reports an approximate average of 300 headings with verified grandparent, 15 with unverified grandparent, and 125 headings with no grandparent in the LCS subject index.

In their recent study of the compatibility between LCSH and subject headings in the catalog at the University of Michigan, Frost and Dede considered how easily conflicting headings might be automatically controlled and what types of corrections would require human intervention. They concluded that the LC Subject Authority File (SAF) and automated authority control systems could effectively be used to control main subject headings and topical and chronological subdivisions but that geographic subdivisions posed more of a problem.

Frost and Dede suggested that lists of free-floating subdivisions from the *LC Subject Cataloging Manual: Subject Headings* might be added to the SAF and a "partsmatching" approach used to authorize more headings than could be authorized by the use of the SAF alone.[6] OSUL has used a list approach and a parts-matching process in a different way to accomplish a similar purpose.

To summarize, HCTL has improved our ability to correct outdated headings automatically by providing us with a means of globally updating subdivisions as well as main headings. It also makes more extensive use of the cross-references in LCS by partially flipping headings when the grandparent portion matches a reference. It has improved our ability to identify problems by automatically grouping headings according to their verification status and relation to other LCS headings.

The HCTL program prevents many variant headings from being added to LCS so that fewer corrections have to be done manually, and correct data is displayed to the public right away. It helps OSUL make the best use of limited human resources by grouping data with common characteristics so that attention can be focused on the areas where it is needed most. This type of programming is particularly valuable when OSUL does

retrospective conversion projects and processes cataloging backlogs—activities that tend to reintroduce outdated, incorrect headings to the file. The specifications of the HCTL program can be modified as LC policy changes and can greatly assist library staff in maintaining the quality and currency of the LCS database.

Table 1. Headings Automatically Corrected and/ or Reported for Review.

Automatic Corrections.	
Subdivision deleted.	30,521
Subdivision replaced.	8,398
Heading partially flipped.	4,429
Reported headings.	
No grandparent*	21,354
Unverified grandparent.	28,934
Verified grandparent.	139,840
City flip.	4,879

*These headings all have at least one subdivision.

REFERENCES

1. Nadine L. Baer and Karl E. Johnson, "The State of Authority," *Information Technology and Libraries* 7:148 (June 1988).
2. Ibid.
3. Lorene E. Ludy and Susan J. Logan, "Integrating Authority Control in an Online Catalog," *Proceedings of the ASIS Annual Meeting* 19:176-78 (1982).
4. For descriptions of the use of the LC name and subject authority tapes at OSUL, see Lorene E. Ludy, "LC Name Authority Tapes Used by Ohio State University Libraries," *Information Technology and Libraries* 3:69-71 (Mar. 1984); and Lorene E. Ludy, "OSU Libraries' Use of Library of Congress Subject Authorities File," *Information Technology and Libraries* 4:155-60 (June 1985).
5. Karen Markey and Diane Vizine-Goetz, *Characteristics of Subject Authority Records in the Machine-Readable Library of Congress Subject Headings* (Dublin, Ohio: OCLC Online Computer Library Center, Aug. 1988).
6. Carolyn O. Frost and Bonnie A. Dede, "Subject Heading Compatibility between LCSH and Catalog Files of a Large Research Library: A Suggested Model for Analysis," *Information Technology and Libraries* 7:288-98 (Sept. 1988).