

# An Analysis of Tables of Contents in Recent English-Language Books

R. Conrad Winke

*A sample of 648 current English-language book publications with Library of Congress cataloging was examined to determine how many have tables of contents suitable for inclusion in bibliographic records. They were also examined to determine the number whose bibliographic records already contain contents notes (MARC field 505) supplied by the Library of Congress, the overall average length of their tables of contents, the levels of complexity or hierarchy of tables of contents, whether the tables of contents were subject-based or author/title based, how many new author names would be added to a bibliographic record that contained an analytic tables of contents note, whether books on certain subjects are more likely than others to include tables of contents, and to determine the proportion of books with usable tables of contents that also have subject indexes which might be usable for enhancing keyword access. Finally, I examined all current bibliographic records produced by the Library of Congress in order to determine how many books in general include subject indexes and how many bibliographic records contain contents notes. It was found that 92.75% of the books examined had tables of contents that could be included in catalog records, with an average length of 67.75 words. Most tables of contents contain one or two levels of hierarchy. Author/title based tables of contents account for 25.62% of the sample pool, with each table containing an average of 15.58 names. Finally, 1.12% of the bibliographic records currently produced by the Library of Congress include contents notes and 53.96% indicate the presence of an index.*

As more people gain the ability to access library collections from remote locations, the luxury of obtaining a minimal citation from a catalog and then browsing the shelves diminishes. Today's user needs more complete information online so that bibliographic selection accurately

can be made remotely and document delivery can follow. However, while the needs and expectations of library users continue to grow, it is already acknowledged that the present generation of online catalogs, and even those catalogs with full Boolean retrieval options, is too lim-

ited in its searching and retrieval capabilities (Mandel 1992). As a reflection of these trends, during the spring of 1992, the Bibliographic Services Study Committee of the Council on Library Resources (CLR) met with representatives from the Library of Congress (LC), experts in Internet access techniques, and online catalog developers. A central theme of that meeting, as defined by Hildreth (1992), was an online catalog (called the E3) that was:

- enhanced with post-Boolean retrieval techniques and improved interface design;
- expanded to include indexing, tables of contents, etc.; and
- extended to serve as a gateway to other collections and resources.

Preliminary investigations showed that retrieval abilities increase when subject-rich information such as transcriptions of tables of contents (TOCs) were added to records and free-text searching was made available. In the present cataloging environment, however, it is more the exception than the rule to include such data. LC greatly restricts the inclusion of contents notes in its cataloging records via its rule interpretations. Furthermore, the "core level" record, currently being touted as the new universally accepted basic cataloging standard, eliminates nearly all notes, including contents notes describing TOCs, with the exception of multipart items with separate titles. Certainly, there are drawbacks to adding such data, such as increased staff workloads and the requirement of more computer storage space, but these drawbacks must be weighed against the benefits. Questions that might be posed include:

- Exactly how much extra effort would be required to enhance more bibliographic records through the addition of contents notes?
- How much extra storage would be required by doing so?
- How much staff time would be needed to accomplish such a task?

It would, of course, be possible simply to implement such a program at a given library and later determine the costs and

storage implications. However, it would be more prudent first to attempt to determine the general nature of TOCs, their length, and the subject areas most suitable for these types of enhancements. In this study, the latter approach is taken in order to provide a foundation for further exploration of this topic.

## LITERATURE REVIEW

Previous research done on contents notes can be divided into two distinct categories: research conducted to prove the benefits of adding contents information to catalog records, and research into the nature and structure of TOCs themselves. This literature review covers both of these categories.

Dwyer (1987, 132) stated that "our clients are requesting not just more and better subject access, but a deeper analysis of the contents of library materials." Matthews, Lawrence, and Ferguson (1983) reported that the ability to search a book's TOC or index is among the special features desired most by respondents to a CLR Online Catalog Evaluation Project questionnaire. Mandel (1983, 6) reviewed proposals for enhancing subject access and noted that "suggestions for adding keywords or uncontrolled terms to the record have included adding words from the table of contents or index." Building on this, Diodato (1986) determined that tables of contents and indexes in books contained terms that patrons used to describe those books. These subject devices could complement traditional library subject headings in providing subject access to books because the additional terms would increase the likelihood that the patron's search term would match information in the bibliographic record.

The assignment of more subject headings per book from a controlled vocabulary would also increase the reader's chance of linking their terms with relevant citations. One frequent justification for adding contents notes to MACHINE Readable Cataloging (MARC) records is the limited number of subject headings attached to records. O'Neill and Aluri

(1981) found about 1.4 subject headings on the average bibliographic record in OCLC, although this number varied somewhat by subject categories and has increased in recent years. Knutson (1991) found an average of 2.6 subject headings per record for social science. He notes (67) that "the very real economic consideration of a known cost [of adding more subject headings] and the not well-known benefits of increasing the level of subject analysis prevents libraries from making a sharp break with tradition in subject cataloging." Despite this, he believes that subject analysis at the chapter level would improve retrieval and increase the likelihood of circulation. However, it is not LC practice or the practice of most libraries to provide chapter level subject access to most books.

Markey and Calhoun (1987) investigated, in part, whether contents notes contributed "subject-rich words" to catalog records. They found that, on average, contents notes added 15.5 unique subject-rich words per record, although they found that only 5% of the more than 35,000 records used in their study actually contained either contents or summary notes.

While no comprehensive study of TOCs has been undertaken until now, there have been a number of partial investigations in the past. Settel and Cochrane (1982) reported on the landmark Subject Access Project (SAP), which provided one of the earliest investigations of the usefulness of adding book tables of contents to library catalogs. Nearly 2,000 books in a number of subject categories in the humanities and social sciences were selected from the University of Toronto Library. The investigators developed rules for selecting subject-rich terms from the TOCs and indexes of books, which were then incorporated into the corresponding bibliographic records. A quota served to restrict the number of terms to be added to each record, resulting in an average of 40. Two separate databases were then set up, one containing the TOCs and index entries, the other containing standard bibliographic records with subject headings. It was found that

twice as many of the relevant items for social science and three times as many relevant items for the humanities were retrieved using the enhanced database. In an earlier article reporting on the SAP, Atherton (1978) reported that of the selected pool of books, only half had usable contents pages, and of these, only some could be deemed to have informative contents pages.

Byrne and Micco (1988) conducted the Enhanced Subject Project (ESP), which was subsequently undertaken by the Australian Defence Force Academy (ADFA). Using the same rules for selecting supplementary terms from the contents and index pages as used in the SAP, librarians at ADFA added terms from the chapter-level headings of contents pages into 653 fields in the bibliographic record for each book. Within six months of application, more than 6,000 books had been processed with enhanced subject access terms.

Both the SAP and the ESP demonstrated that the use of contents terms proved a viable and cost-effective technique for increasing the number of subject access points to the contents of books in library catalogs without seriously compromising precision. The authors also concluded that current cataloging practice, which results in an average of one to two subject headings per record, limits subject access. Both projects showed that recall and user satisfaction were increased when keyword access was provided by adding between 30 to 50 access points per record, paralleling the access provided in online periodical article indexes. The researchers also felt that the controlled vocabulary headings helped provide a grouping function not achieved by keyword terms, and should also continue to be provided in records.

Knutson (1991), however, was unable to link record enhancement to either higher retrieval or to higher circulation, although he felt that enhancing the records with contents notes was a convenient method of increasing the number of terms available for keyword searching, and could increase retrieval potential. In his study, in which 291 col-

lections of essays in the social sciences were enhanced with contents notes and chapter-level subject headings, he found that only 34 of the records (11.7%) originally contained contents notes. One cause for the small gain that Knutson found might be that during the study period, keyword searches accounted for only 2.04% of all searches done in the library's online catalog.

In a similar study, Dillon and Wenzel (1989) constructed a database of nearly 5,000 computer science or electrical engineering book records with abstracts and tables of contents. They found that at all levels of recall the addition of contents information resulted in an improvement of at least 10%, which they considered to be substantial. Concurrently, Dillon and Wenzel found that precision suffered as more content-bearing information was added to records. While the difference in retrieval effectiveness between abstract and TOC information was minimal, the best results were achieved when both were added to the records.

Sellers (1994) looked at whether contents note information would duplicate *Library of Congress Subject Headings (LCSH)* in enhanced records. A sample of 60 records with tables of contents notes from subclasses QP, QE, RC, and SF of *LC Classification* numbers (LCC) were selected. All were found to have terms in their contents notes or titles that duplicated at least 50% of the words in *LCSH* (ignoring differences such as plural or singular form, with form subheadings excluded). For almost half of the records, all of the *LCSH* words were duplicated in the title and contents notes. Sellers was generally discouraged by her results, which were possibly due either to technological restraints or to the search engine in place on the local system. Nevertheless, it was shown that subject access was improved for at least half of this small, topically narrow sample through the inclusion of contents notes.

Michalak (1990) conducted a limited enhancement project at Carnegie Mellon University using both TOCs and abstracts. From a pool of over 1,200 books, 189 (15.2%) were selected as candidates

for enhancement. To be a candidate for enhancement, books had to contain "citable" references in the table of contents, such as those found in anthologies of plays, collections of critical essays written by different authors, and in books with separately-authored chapter titles. Exhibition catalogs with 25 or fewer artists, and books with chapter headings that contained useful contents information otherwise not available in the record, were also enhanced. Eliminated were proceedings of scientific and technical conferences, books for which appropriate keyword access is already provided for in the normal course of description, and books without useful chapter headings (as determined by the cataloger). A retrospective enhancement of more than 3,300 English language collected plays was also undertaken. Of these, 631 records (19%) were deemed eligible and enhanced. Michalak's preliminary results suggested that contents enhancement increased the number of records retrieved by 20% to 30%.

Weintraub and Shimoguchi (1993) examined 375 books in the San Diego State University Library to determine the extent to which information in these books had subject or analytical applications. Books were selected from LCC ranges P-PT, Q-QR, RA-RC, and T-TK. No restrictions were placed on place of publication, date of publication, or language of the work. Books were considered eligible for enhancement if they contained citable works or subject-revealing chapter titles. Books with parts whose titles had little meaning outside the context of the whole, or for which appropriate and sufficient keyword, subject, or name access would be provided routinely, were considered ineligible. Eligible books were divided into two categories that highlighted the type of enhancement that might be used (either citation-based or subject-based). They found that 23% of the books contained discrete content information not already represented in the catalog records. Of those, 52% would call for citation-based enhancements (such as single-author treatises, or multi-author books whose separate chapters were not attrib-

uted to any one contributor) and 48% would call for subject-based enhancements. Nearly 65% required fewer than 25 enhancements, with an average of 8.03 enhancements per book.

Citation-based (or composite) TOCs were also examined by Poulsen (1996). Using sample pools from the collections of two Danish libraries, Poulsen sought to "illustrate the degree to which the proportion of composite works is dependent on library holdings in terms of broad universal holdings versus specialized holdings, subject field, publication language, and the age of the holdings" (136-37). Poulsen, who compared his own results with those of other studies, found that regardless of the various research designs or methods used in examining the sample pools of books, the proportion of books containing composite works fell between 10% and 20%, with the number of articles in the composite works varying from 20 to 30 per book. He also found that books on scientific subjects tended to have more citations than those on the social sciences, and that English-language publications tended to have more citations than publications in other languages.

DeHart and Matthews (1990) explored briefly the structure of TOCs. From a pool of 36 books on information and computer science, they found 14 to have tables of contents with "part" headings in addition to chapter headings. All 36 had chapter titles, with 32 having one chapter sublevel, 15 having two chapter sublevels, 3 having three chapter sublevels, and only 1 having four chapter sublevels. The authors felt that the levels appearing in the TOCs could serve as a weighting device to indicate the importance of various topics within a book.

DeHart and Reitsma (1989) provide a more in-depth investigation into the structure of TOCs, involving a pool of 31 single-work titles on the history of taxation in Great Britain. Publication dates for these books ranged from 1670 to 1981, with over half being from the latter half of the 20th century. Of these, 26 contained TOCs. The authors identified 603 content-indicative terms in the TOCs from these books (an average of 19.5 per book),

although 86% of the chapter titles were deemed meaningless or ambiguous unless understood in conjunction with subject headings or title terminology. Of the 603 content-indicative words, 18 were found to be misleading because they had meanings outside the subject scope of the project. As with the DeHart and Matthews study, levels within the TOCs were identified. Of the 26 books with TOCs, 1 had four chapter sublevels, 3 had three chapter sublevels, 4 had two chapter sublevels, and the remainder had only one sublevel. The authors concluded that a functional relationship exists for enhancing catalogs with contents notes, although they acknowledged that their sample pool was small, narrow in scope, and covered a wide span of publication dates.

Dwyer (1991) also looked into the possibility of enhancing records by selecting 124 titles with LC copy from an approval book shipment. The books were examined to determine the extent of contents notes, especially the number and complexity of "sections" (defined to be primary parts, chapters, essays, or stories with significant title or author/title information excluding prefatory material, bibliographies, indexes, and the like). Dwyer found that the books averaged between 7.5 and 8.5 sections per TOC. Subsections were included in 69 of the books, and sub-subsections in 21 of the books. The TOCs in books with sub-subsections were found to resemble extensive indexes. More detailed TOCs were generally limited to science and technology books. The 55 books with only primary sections averaged a slightly higher 10.9 sections per title, and 19 of these had author/title information on the contents page. Online records for 20 of the 55 books in this final group were enhanced with 505 fields. Dwyer found that this task added approximately nine minutes per title to the cataloging time.

#### LITERATURE SYNTHESIS

A synthesis of the above-cited literature reveals several points. First, users prefer more complete bibliographic records and

desire better access to these records. One way of providing this is through the inclusion of TOCs, which offer not only a deeper analysis of the original material, but also more subject rich keywords per record. Second, recall increases when such vocabulary is added to bibliographic records. However, investigations into the structure and composition of TOCs themselves has been relatively scant, leaving an incomplete and inadequate understanding of TOCs. Sample pools in these studies have been relatively small, ranging in size from 31 to 124 books.

In some cases, only books on certain subjects were investigated. Only one researcher included year of publication as a criterion for inclusion in the sample pool. Furthermore, in none of these papers were TOCs in currently published literature examined to any depth or extent. While some authors had looked at the physical structure of TOCs, none had determined their overall lengths, nor had any investigated materials on a broad range of topics to try to discern whether some subject areas were more likely to include TOCs than others. Finally, none had determined the number of author names that might be added to bibliographic records that include TOCs. These lacunae formed the genesis of the present investigation.

### OBJECTIVES

The goal of this study was to take a detailed snapshot of TOCs as they appear in current English-language book publications. Specifically, the objectives of this study were:

1. To examine current English-language book publications with LC cataloging to determine the percentage of these materials with TOCs suitable for inclusion in bibliographic records
2. To determine the portion of the bibliographic records that are supplied by LC with usable TOCs and already contain contents notes (MARC field 505)
3. To determine the overall average length of TOCs in these books
4. To determine hierarchical complexity of TOCs

5. To determine the proportion of books with usable TOCs that contain subject-based tables and the portion that contain author/title-based tables
6. To determine how many new author names would be added to a record that contained an author/title TOC
7. To determine whether books on certain subjects were more likely than others to include or exclude TOCs
8. To determine the proportion of books with usable TOCs that also have subject indexes that could be potentially mined for further enrichment of keyword access
9. To examine all bibliographic records for books published from 1995 to 1997 with LC cataloging in order to determine: how many indicate the presence of indexes; how many records already contain contents notes; the overall average length of records with and without contents notes; and the average number of words contained in the contents notes themselves

### STUDY POPULATION

For a variety of reasons discussed below, the following limiting guidelines were placed on the study population:

1. Because increased subject retrieval through keyword searching is one of the benefits to be gained from the inclusion of TOCs in cataloging records, this study was limited to English-language books, the assumption being that most North American libraries would be primarily interested in enhancing their keyword indexes with English language words.
2. Books included in the study were limited to those having full-level LC bibliographic records because it was felt that such materials would better mirror the offerings of mainstream presses and thereby most likely be found in the greatest number of North American library collections. Books with LC copy cataloging were also included, as these are materials that LC has collected. Using materials with full LC records

also automatically provided a convenient breakdown by subject via the LCC numbers, which are included in each fully cataloged record. This permitted me to determine whether books on certain subjects were more likely to include usable TOCs than others.

3. Because only recently published books were to be investigated for this study, books with imprint dates earlier than 1995 were excluded from the sample.
4. For practical purposes, the study population was further limited to books collected by the main library of Northwestern University Library (NUL).
5. For obvious reasons, books without a usable TOC were excluded from the study population.

Points 4 and 5 are further elaborated upon below.

#### METHOD

The study took place at NUL during August and September 1997. The main library's collection totaled 3,840,439 volumes as of August 31, 1996 and grew by an average of 27,093 new titles each year between 1992 and 1996. During those same years, NUL processed an annual average of 17,693 titles with LC copy. Being a large research library, its collections are broad and general in nature.

No individual library can truly be regarded as an accurate reflection of the current English-language book publishing industry. LC, being the largest library in the English-speaking world, might be a reasonable approximation, and certainly it catalogs the greatest number of such books. It is assumed that because of its size and the number of titles added each year, NUL's collection might therefore also be used as some indication of the type of monographs currently being published in English and cataloged by LC. However, some limitations to this assumption must be stated.

The Catalog Department in the Main Library at NUL catalogs for the general collection. Separate branch libraries,

each with its own cataloging unit, exist for Law, Medicine, Music, and Transportation. It was expected that these subject areas (LC classifications K, R, M, HE, and parts of T) would be underrepresented in the survey sample. As is the case with any library, NUL's collections are strong in some areas and weak in others. One area of particular strength is *Africana*. However, NUL also collects heavily in history, the social sciences, and political science (LC classifications C, D, E, F, G (partial), H, and J). Other subject areas, such as agriculture (LC classification S) are almost entirely ignored. Being a research library, NUL also collects more heavily in books issued by university presses than in those from commercial publishers, and in books of an academic nature instead of those of a popular nature. Juvenile materials, while collected, were not included in this investigation.

The aim of the study was to create a minimum pool of 600 books with usable TOCs. Books were first sorted by LC classification. Next it was determined whether these books had TOCs that could be analyzed. This two-step process was carried out in order to determine which disciplines (if any) tended to have usable TOCs and which did not. Some examples of books that were determined to not have analyzable TOCs were biographies that did not present subject-like chapter headings, single works of fiction or drama, and some belletristic anthologies (such as poetry) in which the table of contents listed each work as an author-title entry and contained more than 2,500 words, which was assumed to be too lengthy to be included in bibliographic records.

The final candidate pool did not represent a random sample of books received for cataloging in a specified time frame, but rather were the first 601 books received after the study began that met the criteria of having a usable TOC.

Once the final candidate pool had been identified, the bibliographic records for each book were reviewed to determine whether contents notes (MARC field 505) were included. This was done to obtain an idea of how frequently TOCs

were already included in bibliographic records.

Next, a word count for each TOC was determined and the number of levels of hierarchy was recorded. A word was defined as anything with spaces around it. The number of characters in the counted words was not recorded. In accordance with current cataloging practices, chapter numbers, words such as "chapter" and "part" (but see below), and subtitles of chapters or article titles were routinely included in the word counts. Personal author names were also counted, as long as they did not already appear in statements of responsibility elsewhere in the bibliographic record. However, in the spirit of the *Anglo-American Cataloguing Rules*, 2d ed., chapters or articles written by more than three authors were subjected to the rule of three, meaning that only the first author's name was included in the word count. The ignored authors' names were tallied as two words, representing the "[et al.]" that would appear in the contents note in the bibliographic record. Each separate word in an author's name was counted separately (e.g., "A. A. Milne" was counted as three words).

Word counts deeper than chapter level were not carried out. Typically, TOCs with three or more levels of hierarchy included subchapter level analysis. Levels of hierarchy were also ignored unless they contained meaningful words. For example, if a book was divided into parts and chapters, in order for "Part I" to be included in the total word and hierarchy counts, it would have to have a specific name, such as "Part I: Childhood." Other words routinely ignored in the total word counts were chapter names such as "Biography," "Filmography," "Notes," "Acknowledgments," "Contributors," and "Index." Conversely, words such as "Introduction," "Prologue," "Appendix," and "Sources" were included in the total word count if they included more specific titles or if they were integrally built into chapter numbering (e.g. "Chapter 1: Introduction," "Chapter 2: The Early Years," etc.). Finally, in some books that consisted essentially of reprints of previously published articles, the chapter information

contained not only author and title statements, but bibliographic citations as well. These citations were excluded from the final word counts.

It should be noted at this point that there was no attempt to identify the number of "subject-rich" words that might be added to keyword indexes from each TOC. While such information would certainly be useful, such a task would consume more time than was available, due to a number of factors. First, there is the large number of books reviewed in this study and the limited staff available for the project. Second, there is the intellectual problem of determining which words should be regarded as subject-rich, and which should not. Third, in order to obtain an accurate or useful word count, multiple occurrences of a word within the contents note itself, as well as occurrences of that word in other parts of the bibliographic record, would have to be eliminated from a TOC's total final word count. Finally, these subject-rich terms would by their very nature be taken from an uncontrolled vocabulary (that is, the words in the TOCs themselves), and some sort of control would have needed to be established to grapple with the tallying of synonyms or different grammatical manifestations (e.g. singular or plural) of a word occurring in a given record. For all these reasons, it was decided that no effort would be made to tabulate the percentage of the total word count that consisted of potentially subject-rich words.

Each TOC was then reviewed to see whether it was subject-based or author-based. Generally, author-based TOCs are books produced under editorial direction in which a different author writes each chapter, and the names of these authors appear with the chapter title. Subject-based TOCs are generally found in books in which a person or persons is responsible for the intellectual content of the whole book. For author-based TOCs, the total number of new authors' names that the contents note would bring to the catalog record was tallied. Finally, the books were also checked to see whether subject indexes were present, in the hopes of determining whether a correla-



TABLE 1  
DISTRIBUTION OF BOOK SAMPLE BY LC CLASS NUMBER

LC Class	Survey Total	% of Survey Total	Unusable
A	0	0.00	—
B	48	7.41	1
C	5	0.77	1
D	72	11.11	1
E	36	5.56	3
F	12	1.85	—
G	25	3.86	—
H	175	27.01	1
J	46	7.10	—
K	3	0.46	—
L	9	1.39	1
M	1	0.15	—
N	15	2.31	4
P	126	19.44	29
Q	26	4.01	3
R	15	2.31	1
S	2	0.31	—
T	10	1.54	1
U	12	1.85	—
V	0	0.00	—
Z	10	1.54	1
Total	648	99.98	47

tion existed between the presence of a TOC and the presence of an index with potentially usable keywords.

#### STUDY POPULATION COMPARED TO LC

At the same time that books and bibliographic records were being examined, a snapshot was taken of the LC bibliographic records for English-language monographs published in the years 1995 through 1997, for comparison with the survey sample. This was possible because NUL purchases the LC MARC Books Service on tape and mounts the current three to four years' worth of records as a discrete resource file on its local system. All records with usable LC call numbers

and with the language code "eng" were tabulated by class number for comparison with the sample pool. Records with call numbers such as "IN PROCESS" and "MLC" were ignored. This snapshot allowed for an analysis of book cataloging currently being performed at LC by classification number. In addition, the LC records were scanned for the presence or absence of 505 fields and the presence or absence of indexes as determined by the fixed-field index code.

#### FINDINGS

A total of 648 books was examined to produce a pool of 601 titles that included usable TOCs. The distribution by LC

TABLE 2  
LEVELS OF HIERARCHY EXHIBITED IN SAMPLE TOCS

Levels of Hierarchy	No. of Books	%
1	280	46.59
2	258	42.93
3	47	7.82
4	14	2.33
5	1	0.17
6	1	0.17

class number of the books examined, including those deemed unusable, is shown in table 1. (Due to rounding, percentages in the tables do not add to exactly 100%.) None of the 648 books was assigned class A or V. Of those books that were deemed to be lacking in usable TOCs, 29 (61.7%) were in class P. This might well be expected, as works of imagination, such as novels, typically do not contain TOCs. Another category of materials frequently found in class P, biographies of authors, often does not contain meaningful TOCs. It is interesting to note, however, that 601 (92.75%) of the books examined did contain TOCs that could conceivably be included in their corresponding bibliographic records.

Of the 601 TOCs examined, the average number of words in each was 67.75, with a range from 9 to 2,078 words and a median of 81 words. Only 32 (5%) TOCs exceeded 300 words, and in only 2 instances (0.33%) did the number of words exceed 1,000. (As stated above, some of the unusable TOCs were deemed to be so due to excessive length.) Of the 67.75 average words per TOC, not all of these might be considered "subject rich," nor would all of them be new words added to a bibliographic record. Nevertheless, some of them certainly would enhance keyword retrieval and all would give a user a better idea of the contents of a given work without first having to retrieve it from the stacks or order it through interlibrary loan.

Another aspect of TOCs that was examined was structure, or the number of levels of hierarchy. These results are

shown in table 2. Nearly 90% of the TOCs in this study were relatively simple in their structure. The greatest number (46.59%) did not exhibit any hierarchical structure, which is to say that they consisted of a simple list of chapters or parts. TOCs with two levels of hierarchy comprised the next largest group (42.93%). Most of these were arranged into parts and chapters, but some were arranged into chapters and subchapters. The remaining 63 books had between three and six levels of hierarchy in their TOCs. In general, books with four or more levels of hierarchy delved into analysis below the chapter level. Due to their overall length and level of detail, it would in all likelihood be impractical to attempt to include such detailed TOCs in catalog records, which was why the subchapter level analysis portion of these TOCs was omitted from the total word tallies.

As stated above, there are two types of TOCs, subject-based and author/title-based. Of the sample pool, 154 books (25.62%) contained author/title-based TOCs. Each contained an average of 15.58 new authors per book (that is, the authors' names did not already appear in the catalog record). These additional names could provide a user with better and more complete access to the works of a given author via keyword searching, albeit without the added benefit of authority control.

These data were then compared to the snapshot of the LC database, which are presented in table 3. As expected, the NUL sample pool varied predominantly along the lines of NUL's collection

TABLE 3  
 SNAPSHOT OF THE LC DATABASE COMPARED WITH THE SAMPLE TAKEN AT NUL

LC Class	Total LC	%	NUL Survey Total	%	Difference
A	438	0.14	0	0.00	-0.14
B	25,515	8.24	48	7.41	-0.83
C	3,934	1.27	5	0.77	-0.50
D	13,841	4.47	72	11.11	+6.64
E	7,030	2.27	36	5.56	+3.29
F	8,375	2.71	12	1.85	-0.86
G	12,892	4.17	25	3.86	-0.31
H	44,605	14.41	175	27.01	+12.60
J	5,970	1.93	46	7.10	+5.17
K	15,161	4.90	3	0.46	-4.44
L	8,578	2.77	9	1.39	-1.38
M	4,495	1.45	1	0.15	-1.30
N	8,232	2.66	15	2.31	-0.35
P	60,152	19.44	126	19.44	0.00
Q	29,749	9.61	26	4.01	-5.60
R	21,074	6.81	15	2.31	-4.50
S	5,788	1.87	2	0.31	-1.56
T	26,832	8.67	10	1.54	-7.13
U	1,928	0.62	12	1.85	+1.23
V	785	0.25	0	0.00	-0.25
Z	4,113	1.33	10	1.54	+0.21
Total	309,487		648		

strengths and weaknesses, with the strongest variation in classes D, H, and T. Having the MARC tapes available in such a manner allowed for some simple and revealing analysis of these records, both on their own and in comparison to the sample pool.

The LC records were also examined for the presence of contents notes (MARC field 505). It was impossible to determine whether these 505s were constructed from the titles of individual volumes of multivolume sets or from actual TOCs. Of the 309,487 records sampled, only 3,471, or 1.12% were found to contain contents notes. Nearly half of these (1,561) were in class P. In nearly all other classes, less than 1% contained contents

notes. Given this, it can be inferred that LC only rarely includes TOCs in its catalog records. To an extent, this is not surprising, as LC's internal cataloging policies restrict the inclusion of contents notes in bibliographic records for books to six specific situations.

Of the 601 usable books in the NUL sample, 37 records, or 6.17%, included 505 fields. Nearly half fell within two classes, 11 in class H and 7 in class P. It cannot be determined with certainty why the incidence of 505s in the study population in class H population was so much higher than those items held by LC. It should be noted that there was no way to determine whether the copy in the NUL records had come directly from the LC

tapes or from OCLC, leaving open the possibility that some of NUL's records might have been enhanced with TOCs by OCLC member libraries. It should further be noted that this class exhibited the greatest variance between the sample and LC's database. Nevertheless, it would seem safe to surmise that, in general, less than 5% of the books currently cataloged by LC receive contents notes, and that a sizable portion of the records that do receive contents notes are for books assigned to class P.

Information regarding the presence or absence of indexes was also extracted from the LC file. A total of 166,996 (53.96%) records had their fixed field coded to indicate that an index was present. It was impossible to determine from the bibliographic record whether these indexes were subject indexes or name indexes. In the NUL sample, 529 of the 601 (88.17%) had subject indexes. As was the case with TOCs, it was not possible to determine with certainty why the NUL sample differed so greatly from the LC sample. Nevertheless, it is possible that more than half the bibliographic records for English-language books currently cataloged by LC could be enhanced for greater subject access through the inclusion of either keywords from their indexes or even the indexes themselves.

Finally, the overall length of the LC cataloging records and of the contents notes contained therein was tabulated. Records that did not contain contents notes averaged 813.15 characters in length. Those that did include contents notes averaged 1,130.54 characters in length. The average length of the contents notes was 254.02 characters (not 317.39 characters, because some records contained more than one 505 field, meaning that direct subtraction of the two averages of overall length does not yield accurate results), or 38.83 words (a word being defined here as a string of characters separated on either side by spaces, delimiters, or double dashes), with each word having an average length of 6.54 characters. That this word count is nearly half the total of the NUL sample might be the result of some unknown percentage of the LC-created

contents notes not being created from TOCs, but created to reflect the titles of single volumes of multivolume sets, or being partial or incomplete contents.

## CONCLUSION

It has been shown in the literature that the inclusion of TOCs in catalog records increases retrieval potential, both for topical searches and name searches (Settel and Cochrane 1982; Byrne and Micco 1988; Dillon and Wenzel 1989; Michalak 1990). TOCs are also one aspect of description that users demand (Matthews, Lawrence and Ferguson 1983; Mandel 1983). Furthermore, with the growth of the World Wide Web and the ability to access remote databases, it becomes increasingly important for users to be able to learn as much as possible about a given item before incurring interlibrary loan charges. It appears that current cataloging practices are not keeping pace with the growing demands of the times and thus do not result in records capable of serving users as fully as possible. Catalog librarians have not been successful in getting financial and administrative support for such endeavors. The examination of 648 books received in sequential order revealed that 601 had usable TOCs, which suggests that a majority of currently published English-language books might actually include TOCs of use to users. The cataloging rules and the MARC format both provide for the inclusion of TOCs in catalog records, and yet, from the above data, it is readily apparent that TOCs routinely are not being included in current cataloging. One possible reason for this might be that the inclusion of TOCs is regarded as labor intensive and therefore cost prohibitive.

This initial examination suggests, however, that TOCs may not be either complex in structure or lengthy, because those investigated contained on average less than 300 words. A typical typist, entering information at a rate of 60 to 90 words per minute, should be capable of entering this data in no more than five minutes, even given all the formatting details entailed in entering these words into

an enhanced level 505 field in a MARC record. Using optical character recognition to digitize the TOCs and utilizing the cut-and-paste features now available in the graphical interfaces of many online library systems might speed the process further. Lower level staff could also perform these tasks. A study investigating the time required to add TOCs to bibliographic records would make for an informative and enlightening followup to this study.

Most books contain indexes laden with subject-rich vocabulary that could also benefit searching. This, however, is a thornier issue. While the majority of books do contain indexes, they are often quite lengthy. They are also riddled with information, such as page numbers, that would not be of great importance in the bibliographic record. The current MARC format also makes no provision for specifically including book indexes in the variable fields of catalog records. One possibility would be to use the 653 field to include certain subject-rich terms from a book's index. This, however, would add an intellectual element to the process that is not required for the inclusion of TOCs in catalog records, and this element could make the inclusion of book index terms cost prohibitive.

Current trends in cataloging lean toward paring down the amount of description provided in the catalog record. However, providing users with more information in the form of keywords culled from TOCs would greatly benefit users' ability to retrieve relevant materials, and, at the same time, possibly trim unnecessary interlibrary loan costs. Users have always needed both the broad data of cataloging records and the detailed data of the books' TOCs and indexes. As users increasingly do not necessarily have direct access to the books in question due to remote searching, the provision of this detailed data falls upon the librarian's shoulders. TOCs are vital. While further study of a randomly chosen study population is needed to support generalization of the findings presented here to all TOCs, this study suggests we can expect to find that TOCs are relatively straight-

forward and brief. If the nature of TOCs suggested here is confirmed, the next challenge remains to show whether their inclusion in the catalog records is cost effective, and if not, how to make it so.

#### WORKS CITED

- Atherton, P. 1978. Improved subject access to books in on-line library catalogs. In *1st International On-Line Information Meeting, London, 13-15 December 1977*, 131-38. Oxford: Learned Information.
- Byrne, A., and M. Micco. 1988. Improving OPAC subject access: the ADFA experiment. *College & research libraries* 49: 432-41.
- DeHart, F. E., and K. Matthews. 1990. Subject enhancements and OPACs: planning ahead. *Technical services quarterly* 7, no. 4: 35-52.
- DeHart, F. E., and R. Reitsma. 1989. Subject searching and tables of contents in single-work titles. *Technical services quarterly* 7, no. 1: 33-51.
- Dillon, M., and P. Wenzel. 1989. Enhanced bibliographic record retrieval experiments. *OCLC newsletter* 181: 13-14.
- . 1990. Retrieval effectiveness of enhanced bibliographic records. *Library hi tech* 8, no. 3: 43-46.
- Diodato, V. 1986. Tables of contents and book indexes: how well do they match readers' descriptions of books? *Library resources & technical services* 30: 402-12.
- Dwyer, J. 1991. Bibliographic records enhancement: from the drawing board to the catalog screen. *Cataloging & classification quarterly* 13, nos. 3-4: 29-51.
- . 1987. The road to access & the road to entropy. *Library journal* 112, no. 14: 131-36.
- Hildreth, C. R. 1992. Advancing toward the E3 OPAC: the imperative and the path. In *Think tank on the present and future of the online catalog: Proceedings*, 17-38. RASD occasional papers, no. 9. Chicago: ALA.
- Knutson, G. 1991. Subject enhancement: report on an experiment. *College & research libraries* 52: 65-79.
- Mandel, C. A. 1983. Enriching the library catalog record for subject access. *Library resources & technical services* 29: 5-15.

- . 1992. Library catalogs in the 21st century. *ARL* 164: 1-4.
- Markey, K., and K. Calhoun. 1987. Unique words contributed by MARC records with summary and/or contents notes. In *ASIS '87: Proceedings of the 50th ASIS annual meeting*, ed. Ching-chih Chen, 153-62. New York: Greenwood.
- Matthews, J., G. Lawrence, and O. Ferguson. 1983. *Using online catalogs: A nationwide survey, a report of a study sponsored by the Council on Library Resources*. New York: Neal-Schuman.
- Michalak, T. J. 1990. An experiment in enhancing catalog records at Carnegie Mellon University. *Library hi tech* 8, no. 3: 33-41.
- O'Neill, E. T., and R. Aluri. 1981. Library of Congress subject heading patterns in OCLC monographic records. *Library resources & technical services* 25: 63-80.
- Poulsen, C. 1996. Tables of contents in library catalogs: a quantitative examination of analytic catalogs. *Library resources & technical services* 40: 133-40.
- Sellers, M. 1994. Cataloging decision consequences: accessing table of contents in the online catalog. *Colorado libraries* 20: 24-27.
- Settel, B., and P. A. Cochrane. 1982. Augmenting subject descriptions for books in online catalogs. *Database* 5, no. 29: 29-37.
- Weintraub, T. S., and W. Shimoguchi. 1993. Catalog record contents enhancement. *Library resources & technical services* 37: 167-80.

## ***Gaylord's line of archival materials will provide your artifacts with a lifetime of proper storage.***

---

*File Folders*

---

*Archival Laser Labels*

---

*Document Cartons & Storage Cases*

---

*Binders and Enclosures*

---

*Folder Stock and Board*

---

*Environmental Monitoring & Disaster Supplies*

See us on the World Wide Web:

**<http://www.gaylord.com>**



The Trusted Source<sup>®</sup>

Gaylord Bros.  
PO Box 490  
Syracuse, NY 13221-4901

