1987

Unique words contributed by MARC records with summary and/or contents notes

Karen Markey
Karen S Calhoun, *University of Pittsburgh - Main Campus*

Unique Words Contributed
by MARC Records with
Summary and/or Contents Notes

by

Karen Markey

and

Karen Calhoun

Professional titles: Dr. Markey: Assistant Professor, The University of
Michigan, School of Information and Library Studies, 580 Union Dr., Ann Arbor,
MI 48109. Ms. Calhoun: User Documentation Specialist, OCLC Online Computer
Library Center, Inc., 6565 Frantz Rd., Dublin, OH 43017.

Degrees: Dr. Markey: B.A., The Johns Hopkins University; M.L.S., Ph.D.,
Syracuse University. Ms. Calhoun: B.A., Bucknell University; M.S., Drexel
University.

Unique Words Contributed
by MARC Records with
Summary and/or Contents Notes

Abstract

This paper presents new analyses that quantitatively demonstrate the potential of bibliographic contents and summary notes (MARC fields 505 and 520) for enhancing subject access in online catalogs. The findings of previous studies indicate that online catalogs containing augmented subject descriptions drawn from books' tables of contents and/or indexes can improve users' ability to retrieve subject information. Bibliographic contents and summary notes are the rough equivalents of augmented subject descriptions in MARC records. Therefore, it is reasonable to expect that analyses of bibliographic contents and summary notes fields in MARC databases would reveal that these notes contribute a significant proportion of unique subject-rich words to the records in which they occur. Such analyses were begun in the Dewey Decimal Classification (DDC) Online Project.

In the DDC Online Project, DDC information was added to MARC records based on the best match of a record's class number with a class number in the DDC. In one of many analyses of these enhanced MARC records, subject-rich fields were analyzed to determine the average number of unique words (excluding stopwords) contributed <u>successively</u> by subject-rich fields. The sequence of fields in the analysis was: subject headings, title, summary and contents notes, series, sponsor, DDC Relative Index entry, and DDC Schedule captions and notes. Libraries whose records were analyzed in the study and the total number of records per library were: (1) Library of Congress (LC) = 11,865 records; (2) New York State Library (NYSL) = 8,144 records; (3) Public Library of Columbus and Franklin County (PLCFC) = 9,719 records; and (4) University of Illinois at Urbana-Champaign (UI) = 7,316 records.

In this first analysis, subject headings contributed an average of 4.15 unique words per record, while the contents and summary notes fields contributed an average of only 0.42 unique words per record. Unique subject-rich words added to records from DDC information contributed an average of 9.16 unique words per record.

The investigation was extended in a second analysis of the subset of records bearing contents and/or summary notes. The number of bibliographic records with such fields and the percentage of the total number of records per library were: (1) LC = 410 records, 3.46% of LC records; (2) NYSL = 46 records, 0.56% of NYSL records; (3) PLCFC = 201 records, 2.07% of PLCFC records; and (4) UI = 353 records, 4.83% of UI records.

In the second analysis, subject headings contributed an average of 4.84 unique words, while the contents and summary notes fields contributed an average of 15.50 unique words per record. The DDC fields contributed an average of 9.26 unique words per record. For both analyses, figures and tables illustrate the numbers and percentages of unique subject-rich words contributed per library.

Conclusions from these analyses of subject-rich words in MARC records were: (1) summary and/or contents notes did not occur in more than 5% of MARC records (which is in keeping with findings of previous studies); (2) summary and/or contents notes occurred less frequently in local libraries' MARC records than in LC records; (3) when summary and/or contents notes occurred in bibliographic records, these notes contributed a significant proportion of unique subject-rich words; and (4) subject-rich words added to MARC records from the DDC contributed a large proportion of unique subject-rich words regardless of the presence of summary and/or contents notes in the record. The authors recommend that libraries consider integrating either contents and summary notes and/or a library classification into their online catalogs, since these are major contributors of subject-rich words.

Unique Words Contributed
by MARC Records with
Summary and/or Contents Notes

Introduction

This paper presents new analyses that quantitatively demonstrate the potential
of bibliographic contents and summary notes (MARC fields 505 and 520) for
enhancing subject access in online catalogs. Online catalog use studies have
provided evidence that there is much more subject searching of online catalogs
than expected, given the findings of traditional library catalog studies, and
that users have problems with subject searching, particularly in the selection
of subject vocabulary. Few users match their terms with the controlled
vocabulary of the library catalog and/or bother to consult this vocabulary
before or during their online search. And, when asked to identify desired
improvements to online catalogs, users want improvements to enhance subject
searching, especially lists of related terms incorporated into the online
catalog.

Previous research has examined whether catalogs or databases which contain
subject-rich terms from books' tables of contents and/or indexes improve users'
ability to retrieve subject information. Researchers in the Subject Access
Project (SAP) at Syracuse University (Atherton 1978) demonstrated that most
books contain suitable information for producing augmented subject
descriptions. Applying a set of selection rules and a quota system, SAP
researchers created subject descriptions from books' tables of contents and/or
indexes. These subject descriptions were entered in an experimental database
of MARC-like records called BOOKS. The costs of creating, storing, and
searching the enhanced database were not prohibitive, and the subject

descriptions were found to produce a useful vocabulary for online searching. When SAP researchers compared online searches of BOOKS with searches performed against MARC records alone, they discovered that BOOKS searches took less time, retrieved more relevant items, and were more precise. Furthermore, searching BOOKS allowed users to find some items which would have been impossible to locate through LC subject headings or through subject-rich terms in title fields.

Wormell (1981a, 1981b, 1983) carried out theoretical and practical work based on SAP indexing methods. At Lund University Library, she designed and tested a database of records for four hundred books in the field of environmental protection and ecology. Existing machine-readable records for these books were augmented with subject terms and phrases from the books' tables of contents and/or indexes and loaded into a test database which was searchable in free-text mode. She demonstrated that the test database not only permitted detailed subject access, but also was cost-effective, because the indexing terminology was taken from the books themselves. Furthermore, unlike other "deep indexing" methods, SAP indexing did not require special subject knowledge. In another experiment, Wormell applied SAP indexing methods to Swedish Government Official Reports. These reports lacked a subject index, and SAP indexing of the reports' detailed tables of contents, table captions, and figure captions created a means of retrieving information from this previously underused source.

There is a commercially available database, "Superindex," which was created with an indexing approach similar to SAP (Starkoff 1984). The Superindex system, which allows free vocabulary searching, was initially loaded in 1983

with the subject indexes of three thousand carefully selected reference books in the fields of science, engineering, and medicine.

Focusing on nontechnical materials, Hoffman and Magner (1985) pointed out the need for analytic catalogs which facilitate the retrieval of works embedded in literature collections or anthologies. In a sample of books drawn from the collection of the Santa Ana College Library, they found that one book in five was a collection or anthology of some sort. While some embedded literary works are retrievable through the use of specialized reference works such as Granger's Index to Poetry, the authors contend that users would be better served through an enhancement of the catalog with contents information.

Disenchanted with Library of Congress Subject Headings (LCSH), staff at the Australian Defense Force Academy Library implemented an automated library system using a SAP approach (Byrne 1986). Their test file included augmented headings drawn from the books' tables of contents and/or indexes. LCSH was seen as not specific enough, biased toward American terminology, behind current usage in scientific and technical areas, and inadequate for retrieving conference proceedings and collections of articles. When queries were searched against the test file, then against records in the main bibliographic database, and results compared, the library staff found that the augmented headings in the test file provided a valuable alternative to LCSH and a "previously unattained standard of subject access to monographic materials" (Byrne 1986, 87).

Similar results were reported by Posey and Erdmann (1986), who developed a UNIX-based information system at the Purdue University Engineering Library.

Their online file of records for monographs, augmented with edited tables of contents, was found to increase the possibility of patrons' finding specific information. Noting the inadequacy of LCSH for providing subject access to engineering books, the authors listed the advantages of their system: their patrons can enter current engineering terminology, rather than having to learn the controlled vocabulary of LCSH; they can easily assess the relevancy of a book by scanning the display of the table of contents; and articles within books with homogeneous contents, such as conference proceedings, are accessible for the first time.

Data and analyses available from the Dewey Decimal Classification (DDC) Online Project (Markey and Demeyer 1986) are pertinent here. Previous studies have shown that online catalogs containing augmented subject descriptions drawn from books' tables of contents and/or indexes can improve users' ability to retrieve subject information. Contents and summary notes are the rough equivalents of augmented subject descriptions in MARC records. Therefore, it is reasonable to expect that an analysis of contents and summary notes fields in MARC records would reveal that these notes contribute a significant proportion of unique subject-rich words to the records in which they occur. Furthermore, such an analysis would facilitate comparisons of unique subject-rich information contributed by other subject-rich fields in MARC records. Such an analysis was begun in the DDC Online Project.

The DDC Online Project team designed an experimental online catalog, the Dewey Online Catalog (DOC), in which DDC information was incorporated. DOC's database was created from three data sources: (1) machine-readable cataloging (MARC) records in selected subject areas, (2) the nineteenth edition of the DDC

Schedules, and (3) the Relative Index. Four libraries--the Library of Congress (LC), New York State Library (NYSL), Public Library of Columbus and Franklin County (PLCFC), and University of Illinois at Urbana-Champaign (UI)--participated in the DDC Online Project, and there were four Dewey Online Catalogs, i.e., one for each of the four participating libraries. One of the many DDC Online Project analyses was an examination of the fields of the enhanced bibliographic records in the Dewey Online Catalogs (Markey and Demeyer 1986, sec. 9.3). As part of this analysis, the researchers examined the contents of subject-rich fields in bibliographic records to determine the average number of unique words contributed _successively_ by subject-rich fields. Later, the investigation was extended in a second analysis of the subset of records bearing contents and/or summary notes (i.e., MARC fields 505 and/or 520).

The results of the first analysis were provided in the final report of the DDC Online Project to the Council on Library Resources and are also covered in this paper along with the results of the second analysis.

Methodology

Bibliographic records in a MARC or MARC-like format were obtained from the four participating libraries. For the Dewey Online Catalogs, these records were enhanced with DDC Relative Index entries and DDC Schedule captions and notes. Only those fields and subfields whose contents were used for indexing, display, and/or processing were kept in bibliographic records in the Dewey Online Catalogs. Of these, seven fields--subject headings, title, bibliographic

notes, series, sponsor, DDC Relative Index entry, and DDC Schedule captions and
notes (collapsed into a single category)--were considered subject-rich fields.
Table 1 lists, for each library, the subject-rich fields in DOC bibliographic
records, the average number of these fields per bibliographic record, and the
percentage of records without these fields.

Table 1. Average Numbers and Percentages of Subject-Rich Fields in Bibliographic Records

Field	LC Avg. per Record	Without Field % in Database	NYSL Avg. per Record	Without Field % in Database	PLCFC Avg. per Record	Without Field % in Database	UI Avg. per Record	Without Field % in Database
DDC Relative Index Entry	2.05	20.48	1.48	11.70	1.51	23.27	3.48	1.34
DDC Caption	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
DDC Notes	1.43	27.70	0.33	70.38	0.40	62.13	1.19	22.12
LC Subject	2.26	0.04	1.14	35.98	1.41	2.76	1.46	0.70
Sponsor	0.04	96.47	0.29	76.61	0.00	100.00	0.04	96.25
Title	1.00	0.00	1.00	0.00	1.00	0.00	1.00	0.00
Series	0.33	68.23	0.01	98.67	0.02	98.02	0.36	64.80
Bibliographic Notes	0.03	96.54	0.01	99.44	0.02	97.93	0.05	95.17

Number of records = 11,865 (LC); 8,144 (NYSL); 9,719 (PLCFC); 7,316 (UI)
Number of fields = 235,050 (LC); 105,320 (NYSL); 121,982 (PLCFC); 105,529 (UI)

The NYSL database had the largest percentage of records (35.98%) without
subject headings. The LC database had the most subject headings per
record--2.26. The source fields for the line in table 1 marked "Bibliographic
Notes" were MARC fields 505 (Contents Note, Formatted) and 520 (Summary,
Abstract, Annotation, Scope Note, etc.). Most bibliographic records lacked

these fields; for instance, over 96% of the LC records contained neither a 505 nor a 520 note. The reasons for this will be taken up shortly.

Two analyses of the subject-rich fields in DOC bibliographic records were performed and compared:

● The contributions of unique words in subject-rich fields of all bibliographic records

● The contributions of unique words in subject-rich fields of bibliographic records bearing bibliographic notes fields (i.e., 505 = Contents, 520 = Summary)

First, the contents of the subject-rich fields in all bibliographic records from each of the four libraries were analyzed to determine the average number of unique words (excluding stopwords) contributed successively by the subject-rich fields listed in table 1. This analysis required the calculation of the number of words per field. These words were then compared to one another, and duplicate words per field were discarded. Detailed results of this preliminary analysis are presented elsewhere (Markey and Demeyer 1986, appendix O). It is sufficient to note here that 36% of the words in Relative Index entries were duplicate words, and 23% of the words in subject headings were duplicate words. The title and sponsor fields resulted in the smallest number of duplicate words. The percentage of duplicate words in the bibliographic notes fields was only 17%. This analysis continued with the calculation of the number of unique words contributed successively by the seven subject-rich fields in bibliographic records. The sequence of the analysis

was: subject headings, title, bibliographic notes, series, sponsor, Relative Index entry from the DDC, and DDC Schedule captions and notes.

Second, bibliographic records bearing summary and/or contents notes (i.e., 505 and/or 520 fields) were analyzed separately. The methodology for the second analysis was the same.

The number of records processed in the first analysis equaled the number of records in the experimental online catalog database for each participating library. This number of records included records bearing bibliographic notes fields:

1. LC = 11,865 records

2. NYSL = 8,144 records

3. PLCFC = 9,719 records

4. UI = 7,316 records

The second analysis focused on the subset of records bearing bibliographic notes fields (i.e., 505 and/or 520 fields). These are listed below along with the percentage of records in each database with these fields:

1. LC = 410 records; 3.46% of records in database

2. NYSL = 46 records; 0.56% of records in database

3. PLCFC = 201 records; 2.07% of records in database

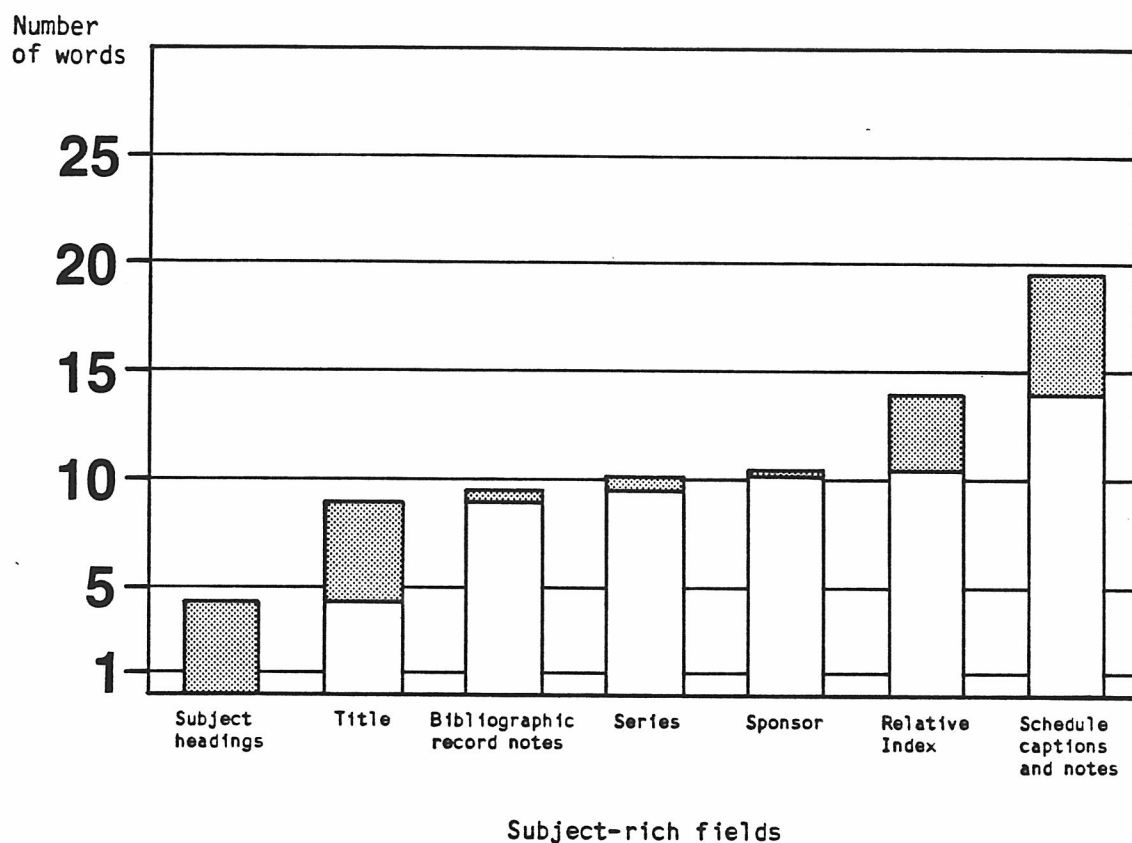4. UI = 353 records; 4.83% of records in database

Results

Table 2 lists the number and average number of unique words contributed by subject-rich fields in all bibliographic records. The third column of table 2 lists the average number of unique words in bibliographic notes fields, which ranges from 0.05 unique words in NYSL records to 0.70 unique words in LC records. Figure 1 is a bar graph showing the successive contributions of unique words in 37,044 bibliographic records, that is, in all the databases combined. The bibliographic notes field contributed 0.42 unique words per record.

Table 2. Successive Contributions of Unique Words in Subject-rich Fields of Bibliographic Records

	Subject Headings	Title	Bibliographic Notes	Series	Sponsor	Relative Index Entries	Schedule Captions and Notes	Total
LC								
Total unique words	73,925	60,234	8,342	16,300	1,109	50,244	120,889	331,043
Average	6.23	5.08	0.70	1.38	0.09	4.23	10.19	27.90
Percentage of total	22%	18%	3%	5%	0.3%	15%	37%	100.3%*
NYSL								
Total unique words	28,021	54,095	429	392	4,924	25,806	25,258	138.925
Average	3.44	6.64	0.05	0.05	0.61	3.17	3.10	17.06
Percentage of total	20%	39%	0.3%	0.3%	4%	19%	18%	100.6%
PLCFC								
Total unique words	30,225	31,279	2,778	512	0	21,768	42,978	129,540
Average	3.11	3.22	0.29	0.05	0	2.24	4.42	13.33
Percentage of total	24%	24%	2%	0.4%	0	17%	33%	100.4%
UI								
Total unique words	21,750	36,353	4,118	10,509	600	29,594	22,633	125,557
Average	2.97	4.97	0.56	1.44	0.08	4.05	3.09	17.16
Percentage of total	17%	29%	3%	8%	0.5%	24%	18%	99.5%
Grand total of unique words	153,921	181,961	15,667	27,713	6,633	127,412	211,758	725,065
Average	4.15	4.91	0.42	0.75	0.18	3.44	5.72	19.57
Percentage of grand total	21%	25%	2%	4%	1%	18%	29%	100%

*Some totals are slightly more or less than 100% due to rounding.

Number
of words

Subject-rich fields

NOTE: Darkened portion of bar represents the number of unique words contributed by this
field. White portion of bar represents number of unique words already contributed by
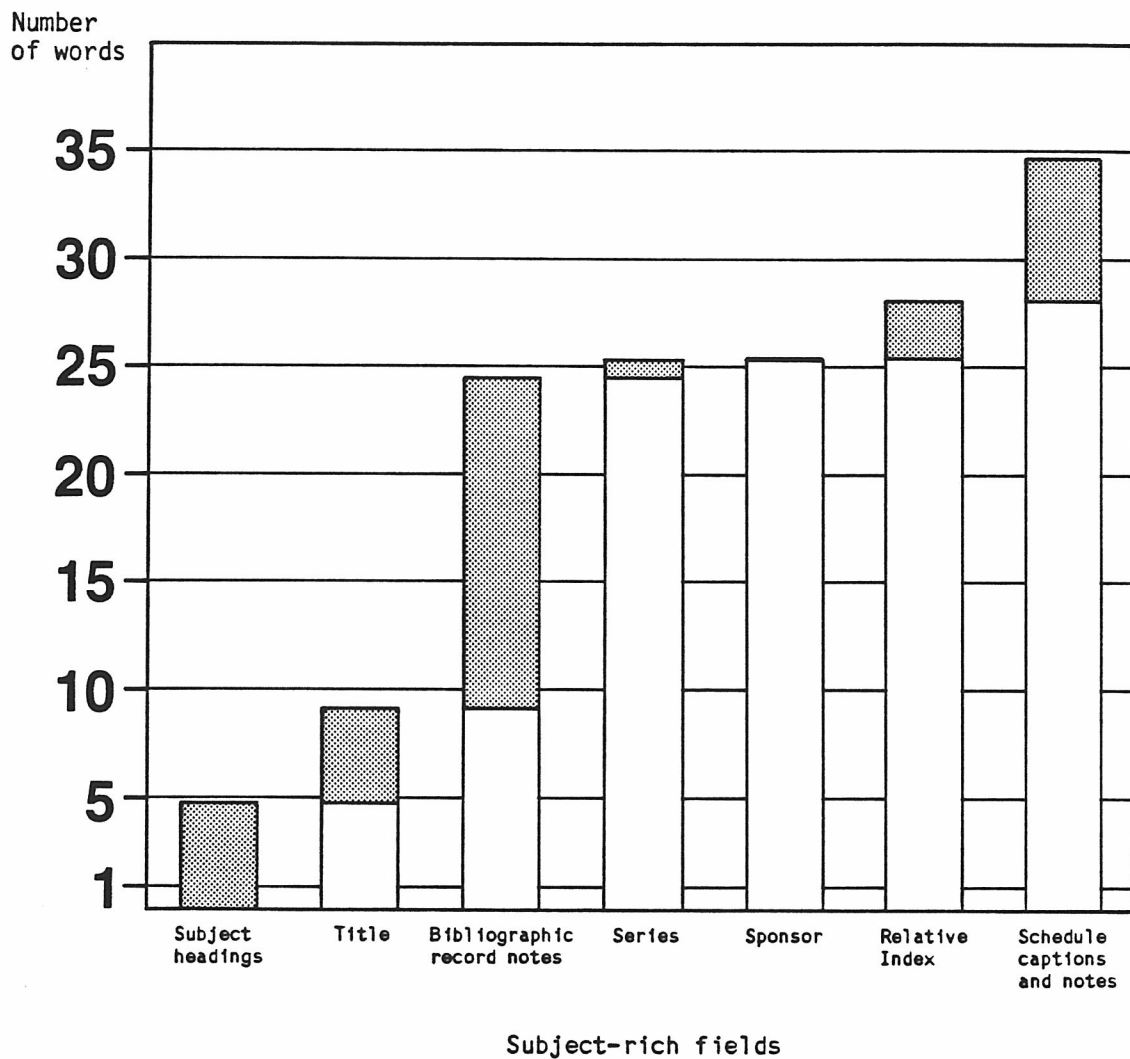fields listed to the left of the bar.

Fig. 1. **Successive contributions of unique words in subject-rich fields
in 37,044 bibliographic records**

The third column of table 3 lists the number and average number of unique words contributed by subject-rich fields in bibliographic records bearing notes fields (i.e., summary and/or contents notes). The average number of unique words in records with a bibliographic notes field ranges from 9.33 unique words in NYSL records to 20.35 unique words in LC records. In this subset, between 36% and 51% of the unique subject-rich words were contributed by the notes field. Figure 2 is a bar graph showing the successive contributions of unique words per field in the 1,010 records bearing bibliographic notes fields. In this subset, the bibliographic notes field contributed 15.50 unique words per record.

Table 3. Successive Contributions of Unique Words in Subject-rich Fields of Bibliographic Records with Notes Fields

	Subject Headings	Title	Bibliographic Notes	Series	Sponsor	Relative Index Entries	Schedule Captions and Notes	Total
LC								
Total unique words	2,921	1,823	8,342	524	63	1,637	4,402	19,712
Average	7.12	4.45	20.35	1.28	0.15	3.99	10.74	48.08
Percentage of total	15%	9%	42%	3%	0.3%	8%	22.7%	100%
NYSL								
Total unique words	285	175	429	26	0	128	137	1,180
Average	6.20	3.80	9.33	0.57	0	2.78	2.98	25.65
Percentage of total	24%	15%	36%	2%	0	11%	12%	100%
PLCFC								
Total unique words	670	587	2,778	38	0	413	993	5,479
Average	3.33	2.92	13.82	0.19	0	2.05	4.94	27.26
Percentage of total	12%	11%	51%	1%	0	7%	18%	100%
UI								
Total unique words	1,014	1,674	4,118	435	11	1,056	597	8,905
Average	2.87	4.74	11.67	1.23	0.03	2.99	1.69	25.23
Percentage of total	11%	19%	46%	5%	0.1%	11.9%	7%	100%
Grand total of unique words	4,890	4,257	15,658	1,023	74	3,234	6,119	35,255
Average	4.84	4.21	15.50	1.01	0.07	3.20	6.06	34.91
Percentage of grand total	14%	12%	45%	3%	0.2%	9%	17%	100.2%

Number of bibliographic records in analysis:
LC = 410; NYSL = 46; PLCFC = 201; UI = 353

Fig. 2. Successive contributions of unique words by subject-rich fields in 1,010 bibliographic records with a summary and/or contents note

NOTE: Darkened portion of bar represents the number of unique words contributed by this field. White portion of bar represents number of unique words already contributed by fields listed to the left of the bar.

The two analyses were repeated, but the subject heading fields were not included, because there are libraries abroad whose records do not contain subject headings. Table 4 lists the number and average number of unique words contributed by subject-rich fields in all bibliographic records. There were fewer subject-rich words per record when subject headings were dropped from the analysis. Most unique words came from the title, Relative Index entries, and Schedule captions and notes. Bibliographic notes fields contributed an average number of unique words ranging from 0.06 to 0.74 unique words.

Table 4. Successive Contributions of Unique Words in Subject-rich Fields of Bibliographic Records (without Subject Headings)

	Title	Bibliographic Notes	Series	Sponsor	Relative Index Entries	Schedule Captions and Notes	Total
LC							
Total unique words	80,792	8,787	17,842	1,592	57,443	141,182	307,638
Average	6.81	0.74	1.50	0.14	4.84	11.90	25.93
Percentage of total	26%	3%	6%	1%	18%	46%	100%
NYSL							
Total unique words	64,614	448	431	5,317	30,831	25,943	127,584
Average	7.93	0.06	0.05	0.65	3.79	3.18	15.66
Percentage of total	51%	0.4%	0.3%	4%	24%	20%	99.7%
PLCFC							
Total unique words	41,128	2,850	545	0	25,020	43,818	113,361
Average	4.23	0.29	0.06	0	2.57	4.51	11.66
Percentage of total	36%	2.5%	0.5%	0	22%	39%	100%
UI							
Total unique words	42,005	4,175	10,680	622	35,524	23,560	116,566
Average	5.74	0.57	1.46	0.09	4.85	3.22	15.93
Percentage of total	36%	4%	9%	1%	30%	20%	100%
Grand total of unique words	228,539	16,312	29,585	7,732	150,274	250,238	682,680
Average	6.17	0.44	0.80	0.21	4.06	6.75	18.43
Percentage of grand total	34%	2%	4%	1%	22%	37%	100%

Table 5 lists the number and average number of unique words contributed by subject-rich fields in the subset of records bearing notes fields. In this subset, bibliographic notes fields contributed an average number of unique words ranging from 9.74 unique words to 21.05 unique words per record. The percentage of unique words contributed by the bibliographic notes fields ranged from 42% to 55%.

Table 5. Successive Contributions of Unique Words in Subject-rich Fields of Bibliographic Records with Notes Fields (without Subject Headings)

	Title	Bibliographic Notes	Series	Sponsor	Relative Index Entries	Schedule Captions and Notes	Total
LC							
Total unique words	2,481	8,631	570	67	1,780	4,544	18,073
Average	6.05	21.05	1.39	0.16	4.34	11.08	44.08
Percentage of total	14%	48%	3%	0.3%	10%	25%	100.3%
NYSL							
Total unique words	259	448	29	0	176	142	1,054
Average	5.63	9.74	0.63	0	3.83	3.09	22.91
Percentage of total	25%	42%	3%	0%	17%	13%	100%
PLCFC							
Total unique words	763	2,850	40	0	469	1,007	5,129
Average	3.80	14.18	0.20	0	2.33	5.01	25.52
Percentage of total	15%	55%	1%	0	9%	20%	100%
UI							
Total unique words	1,832	4,175	442	11	1,350	638	8,448
Average	5.19	11.83	1.25	0.03	3.82	1.81	23.93
Percentage of total	22%	49%	5%	0.1%	16%	8%	100.1%
Grand total of unique words	5,333	16,095	1,081	78	3,775	6,321	32,683
Average	5.28	15.94	1.07	0.08	3.74	6.26	32.36
Percentage of grand total	16%	49%	3%	0.2%	12%	19%	99.5%

Number of bibliographic records in analysis:
LC = 410; NYSL = 46; PLCFC = 201; UI = 353

Examples of bibliographic notes from LC, NYSL, PLCFC, and UI records follow. Both 505/contents and 520/summary notes are labeled "CONTENTS" in these examples.

LC:

TITLE   Agenda for Britain.

LC SUBJECT   Great Britain--Economic policy--1945- --Addresses, essays, lectures.

DEWEY SUBJECT   Economic situation and conditions. Geographical treatment (Economic geography).

CONTENTS   1. Micro policy choices for the 80's--2. Macro policy choices for the 80's.

NYSL:

TITLE   New York: a chronological & documentary history, 1524-1970.

LC SUBJECT   New York (City)--History

                New York (City)--History--Chronology

                New York (City)--History--Sources

DEWEY SUBJECT   History of New York. History of New York City Borough of Manhattan (Manhattan Island, New York County).

CONTENTS   Includes a chronology of historical events in the history of New York City from 1524 to 1970 and a selection of pertinent documents.

TITLE   Revolutionary War veterans buried in Columbia County, New York.

LC SUBJECT   Registers of births, etc.--Columbia County, N.Y.

                Columbia County, N.Y.--Genealogy

                United States--History--Revolution, 1775-1783--Registers lists,

                etc.

DEWEY SUBJECT   History of Other southeastern counties. History of Columbia

                County.

CONTENTS   SET: For holdings consult librarian.

PLCFC:

TITLE   The modern American theater; a collection of critical essays

LC SUBJECT   Theater--United States

                Drama--20th century--Addresses, essays, lectures

DEWEY SUBJECT   Historical and geographical treatment. Geographical and

                personal treatment.

CONTENTS   The attempted dance: a discussion of the modern theater, by A.

                B. Kernan.--American blues: the plays of Arthur Miller and

                Tennessee Williams, by K. Tynan.--The world of Thornton Wilder,

                by T. Guthrie.--The comedy of Thornton Wilder, by T.

                Bogard.--The men-taming women of William Inge: The dark at the

                top of the stairs, by R. Brustein--The theater of Edward Albee,

                by L. Baxandall.--What's the matter with Edward Albee? By T.

                Driver.--Broadway, by F. Fergusson.--Off-Broadway: editor's

                note.--The pass-the-hat theater circuit, by E.E.

Lester.--Happenings in the New York Scene, by A. Kaprow.--The
juggernaut of production, by G. Rogoff.--The drama is coming
now, by R. Gilman.--Which theater is the absurd one? By E.
Albee.--Bibliography (p. 181-183).

UI:

TITLE   Differential geometry in the large : seminar lectures, New York
               University, 1946 and Stanford University, 1956.

LC SUBJECT   Geometry, Differential
               Global differential geometry

DEWEY SUBJECT Analytic geometries. Differential and integral geometry.

CONTENTS   Selected topics in geometry : New York University, 1946 / notes
               by Peter Lax -- Differential geometry in the large : Stanford
               University, 1956 / notes by J.W. Gray.

Notes fields in LC and UI records usually detailed a book's table of contents.
NYSL and PLCFC bibliographic notes were a two- to three-line summary,
availability statement, or a book's table of contents. A number of records
with a contents note had subject heading subdivisions such as "--Congresses,"
"--Addresses, essays, lectures," "--Sources," "--Juvenile literature," and
"--Bibliography."

MARC Formats for Bibliographic Data (Library of Congress 1980) indicates that
field 505 is used for formatted (or formal) contents notes, while field 520
contains an informal note that gives information pertaining to the scope and

general contents of the materials (thus, putting an availability statement in field 505 or 520 is inconsistent with the definitions of these fields). AACR2 (American Library Association 1978) directs catalogers to supply contents and summary notes where they would be appropriate. However, in the OCLC system, catalogers are required to use field 505 only in full-level cataloging records for certain nonbook materials. In records for books, the use of field 505 is optional, and is generally restricted to records for multi-volume sets in which each volume has its own title, or to records for collections of separately titled works. The most current LC interpretation of the AACR2 rule pertaining to contents notes for books (rule 2.7B18) spells out specific instructions for the use of formal contents notes (Library of Congress 1984). It is not uncommon for local cataloging departments to restrict the use of these notes to even narrower categories than LC does.

Field 520 is valid in MARC records for all types of library materials, but in the OCLC system, its use is optional in full-level cataloging records. In records for books, LC catalogers regularly use field 520 only for annotations that appear on records describing children's books. In accordance with the latest LC interpretation of the rule applicable to summary notes for books (AACR2 rule 2.7B17), LC catalogers include 520 notes in very few other cases. Catalogers generally use field 520 more often in records for nonbook materials (such as films), where users cannot easily browse through the item to determine its contents, or where the title is uninformative. In large collections of MARC records, only a small percentage describe nonbook materials; the great majority of the records describe books, and, as this discussion has shown, current cataloging standards rarely direct catalogers to supply contents and

summary notes for books. It is a logical outcome that the DOC databases analyzed in this paper contained relatively few of these notes.

Conclusions

From these analyses of bibliographic records with bibliographic contents and/or summary notes fields, it can be concluded that:

1. The notes field does not occur in more than 5% of bibliographic records. This is in keeping with the findings of previous studies and a logical outcome in terms of current cataloging standards and practices.

2. The notes field may occur less frequently in OCLC member-contributed records than in LC/MARC records.

3. When the notes field occurs in LC/MARC and OCLC member-contributed records, it contributes a large proportion of unique subject-rich words, especially in LC/MARC records.

4. Subject-rich words added to MARC records from the DDC contribute a large proportion of unique subject-rich words regardless of the presence of notes in the record. Notes fields in LC/MARC and OCLC member-contributed records vary in completeness and content. Some detail tables of contents, some provide a summary or annotation, and other do not contain subject-rich words as in the case of availability statements.

The results presented in this paper confirm the potential of contents and summary notes for enhancing subject access in online catalogs. Such notes contribute a large proportion of unique subject-rich words to bibliographic records in which they occur. These findings support the view that augmented subject descriptions created according to SAP techniques can enhance users' ability to locate subject information in online catalogs. (Options for the display of such descriptions in online catalogs, although discussed in some of the research cited here, are not treated in this paper, which focuses on retrieval issues only.)

Experimentation with SAP indexing methods in test databases suggests that users find subject information more easily in online catalogs which include subject-rich terms from books' tables of contents and/or indexes. Researchers testing the SAP approach have found it cost-effective because it uses the terminology of the books themselves and because indexing does not require special subject knowledge. Experience with new optical scanning technology such as the OPTIRAM/LIBPAC system (Harrison 1985) may demonstrate cost-effective techniques for converting book indexes and tables of contents into machine-readable form.

Libraries, particularly those whose records do not contain subject headings, should consider integrating either contents and summary notes and/or a library classification schedule and index into their online catalogs, since these are major contributors of subject-rich words. The MARC fields (i.e., 505 and 520 fields) used to enter contents and summary notes in MARC databases now occur in few bibliographic records, but given the right kind of library cooperation and guidelines, they could be employed more widely to augment subject descriptions of library materials.

References

American Library Association. 1978. Anglo-American Cataloguing Rules. 2d ed.
Chicago: American Library Association.

Atherton, Pauline. 1978. Books are for use: Final report on the Subject
Access Project to the Council on Library Resources. Syracuse, NY: Syracuse
University.

Byrne, Alex. 1986. Life wasn't meant to be whimsical: Painless subject
augmentation. Australasian College Libraries 4(2): 83-90.

Harrison, Martin. 1985. Retrospective conversion of card catalogues into full
MARC format using sophisticated computer-controlled visual imaging
techniques. Program 19(3): 213-230.

Hoffman, Herbert H., and Jereul L. Magner. 1985. Future outlook: Better
retrieval through analytic catalogs. Journal of Academic Librarianship 11
(2): 151-53.

Library of Congress. Automated Systems Office. 1980. MARC formats for
bibliographic data. Washington, D.C.: Library of Congress.

Library of Congress. Processing Services. 1984. Cataloging Service Bulletin
25(Summer): 40-43; 24(Spring): 11.

Markey, Karen, and Anh Demeyer. 1986. <u>Dewey Decimal Classification Online</u> <u>Project: Evaluation of a library schedule and index integrated into the</u> <u>subject searching capabilities of an online catalog</u>. OCLC Research Report series, no. OCLC/OPR/RR-86/1. Dublin, OH: OCLC.

Posey, Edwin D., and Charlotte A. Erdmann. 1986. An online UNIX-based engineering library catalog: Purdue University Engineering Library. <u>Science and Technology Libraries</u> 6(4): 31-43.

Starkoff, B.J. 1984. Superindex: Today and tomorrow. Paper presented at annual meeting of the Society for Scholarly Publishing, 31 May, 1984.

Wormell, Irene. 1981a. SAP: A new way to produce subject descriptions of books. <u>Journal of Information Science</u> 3: 39-43.

_____. 1981b. Subject Access Project: The use of book indexes for subject retrieval system in libraries. <u>International Forum on Information</u> <u>and Documentation</u> 6(4): 24-28.

_____. 1983. Factual data retrieval according to SAP-technique. <u>International Forum on Information and Documentation</u> 8(3): 13-15.