

Tables of Contents in Library Catalogs: A Quantitative Examination of Analytic Catalogs

Claus Poulsen

Easy access to tables of contents from vendors and the technological development of optical character reading have actualized access to articles in books via tables of contents in library catalogs. From earlier studies we know that analytic book catalogs can provide access to up to 600% more works than the traditional catalog by simply adding analytics for works in composite works to the catalog. In this study we examine the proportion of composite works and the number of articles in these books in two different university libraries. The influences of library type, publication language, subject field, and date of publication are examined, and the results are compared to previous studies. The proportion of composite works is between 10% and 20%. The number of articles in the composite works varies from 20 to 30 articles per book—highest for the sciences and the English-language publications and lowest for the social sciences.

Access to articles from composite works has always been valued by users. Whereas journal articles are well indexed through national and international services, it is not easy for the user to get access to articles in books. Some indexes make analysis of some composite works. Some vendors offer access to tables of contents from many books. But since the publication patterns for books are geographically, institutionally, and culturally much more heterogenous than for journals, the value of these indexes is limited.

A natural improvement of library ser-

vices, therefore, would be access to articles in books via the online catalog. The possibility of retrieving electronic tables of contents from several vendors and the development of optical character reading technologies make it realistic to implement such a service.

To grasp the potential of such an improvement we will study the number of articles from composite works that would be accessible in such an analytic catalog. In other words, how many extra citable references for works are added to the online catalog via articles in books?

CLAUS POULSEN is Research Librarian, Roskilde University Library, Denmark (e-mail: claus@emma.ruc.dk). The author wishes to thank Anni Lenschau-Teglers and Nils Bredsdorff for kind assistance during the data collection and Tamara Weintraub for helpful discussions on the subject. Manuscript received June 16, 1995; revised December 20, 1995; accepted for publication January 13, 1996.

ENHANCED CATALOG RECORDS

Traditional library catalogs do not offer access to articles in books. The bibliographic records for composite works do not contain access points for the authors and titles from the articles that make up the book. In an analytical catalog the records are so enhanced, which provides direct access to author and title citations for articles in books.

For years studies have indicated the desirability of improving access to bibliographic records in library catalogs through the provision of contents notes, tables of contents, abstracts, and other methods (Van Orden 1990). All studies show an increased recall in subject searching when contents notes are added.

The traditional question is whether diminished precision from the addition of contents information can invalidate the increased recall. Some studies show that precision does not suffer from this addition of contents information (Atherton et al. 1978; Cochrane 1985; Byrne and Micco 1988; Beatty 1991). Other studies show decreasing precision (Dillon and Wenzel 1990; Knutson 1991). For the time being, apparently, no unambiguous answers exist to the question put in such a general way.

Another substantial question is whether retrieving a large number of contents notes would cause information overload. By increasing the natural text volume dramatically, the number of records retrieved in a topical search will increase (Lancaster, Elliker, and Connell 1989). This was demonstrated by Byrne and Micco in a pilot study and later reaffirmed by Beatty in a full-scale study (Byrne and Micco 1988; Beatty 1991). By adding content-bearing words from tables of contents to the catalog, they reported a 300% increase in the number of items found. One consequence could be that improved analytic access might render the most popular strategy to overcome information overload no longer useful—topical searching using title words in books (Larson 1991).

Apart from the possibility of information overload, this problem will become

increasingly important as the volume of text in library catalogs increases. Consequently, we have to develop a strategy for coping with online catalogs containing increasing numbers of records, which in turn have been enhanced with an increasing amount of description of the contents of the books. One proposed strategy is to separate the enhanced descriptions and offer the choice between traditional subject searching in classification, keywords, and titles or searching where the detailed subject descriptions from the enhanced records are added (Beatty 1991). Another—more radical—proposal has been put forward independently by Lancaster et al. (1991) and Poulsen (1990). In both papers, the authors suggest improving subject access in library catalogs by an enhanced subject description combined with a reduction of the number of records meant for this enhancement. The selected records for this subcatalog represent the encyclopedic or the survey literature.

QUANTITATIVE RESULTS FROM PREVIOUS STUDIES OF ANALYTIC BOOK CATALOGS

Analytic book catalogs are catalogs enhanced by authors and titles from articles in composite works. Composite works are books with two or more distinct works by the same or different authors. Consequently, a composite work consists of two or more citable works. Edited works, anthologies, and conference proceedings are examples.

To show the potential for enhancing the online catalog in this way requires an examination of the number of composite works in the library and the number of articles or citable works in these books.

Hoffman and Magner (1985), working with a sample of 4,094 books in the Santa Ana College Library, found that 21.3% were composite works, containing an average of 31.2 articles or citable works each. In other words, an analytical catalog could add access to 600% more citable works than the traditional library catalog!

Their result is supported by a sample of 446 items in the 1982 cumulation of the American Book Publishing Record show-

ing 22.2% multiple-work documents or composite works (Hoffman and Magner 1985, 152).

Although many authors have studied catalogs with enhanced contents information since the Hoffman and Magner study, they rarely specified the number of composite works and citable works in the catalog (Byrne and Micco 1988; Beatty 1991; Michalak 1990). Only Weintraub and Shimoguchi (1993) did so in their study of catalog record enhancement. They reported that 44 out of a sample of 375 books from the San Diego University Library were composite works. The selection was made from a broad subject field, dominated by language and literature and medicine (see table 1). Further, they reported 1,978 citable references in these 44 books. A recalculation of their statistics shows that 12% \pm 3% of the books are composite works with an average of 4,522 articles in these composite works. The confidence intervals are at the 95% interval. If we calculate the 95% confidence interval in the Hoffman and Magner study we get 21.3% \pm 1.3% composite works to

be compared to 12% \pm 3% composite works in the Weintraub and Shimoguchi study.

The difference between these results is significant. It might be attributed to the different subject fields of the examined books. Weintraub and Shimoguchi's study indicates a difference in the number of articles in composite works between different subject fields (Weintraub and Shimoguchi 1993, 176). The difference can also be due to differences in library type, acquisition policy, and the age of the collection. And finally the two studies show methodological differences.

The explicitly described criteria for selection of composite works show a difference. Weintraub and Shimoguchi use the information in the table of contents exclusively, whereas Hoffman and Magner could increase the number of "multiwork documents" by looking inside the book (Weintraub and Shimoguchi 1993, 170, 172-73; Hoffman and Magner 1985, 152).

These results raise new questions. Can we predict the number of composite works and of articles in these books? Are

TABLE 1
COMPOSITE WORKS IN LIBRARY CATALOGS

Library	Selection	No. Books Examined	Composite Works (%)	Mean No. Articles in Composite Works
Santa Ana College*	All books	4,098	21.3 \pm 1.3 [†]	31
San Diego State Univ. Lib. [‡]	P-PT, Q-QR, RA-RC, T-TK in LC class. [‡]	375	12 \pm 3	45 \pm 22 [#]
Nat. Lib. Educ.	PY<1990	718	16 \pm 3	20 \approx 4
Nat. Lib. Educ.	PY \geq 1990	210	17 \pm 5	17 \approx 5
Nat. Lib. Educ.	English language and PY=1980:1990	496	15 \pm 3	29 \approx 9
Roskilde Univ.	Social sciences and PY=1988:1992	887	24 \pm 3	17 \approx 3
Roskilde Univ.	Science and PY=1988:1992	698	18 \pm 3	28 \approx 5

All \pm intervals are 95% confidence intervals.

* Hoffmann and Magner 1985.

[†] See appendix.

[‡] Weintraub and Shimoguchi 1993.

[§] The contributions are dominated by language and literature (60%) and medicine (24%) with no or negligible contributions from social sciences, science, and technology.

^{||} See appendix.

[#] See appendix.

there real differences between library collections? Are there differences between collections in broad universal libraries versus specialized libraries? Do collections vary by subject fields, publication languages, and the age of the library holdings? Depending on the answers, are there 10% or 20% composite works in a given library and do these books contain 20 or 50 articles each? Will the proportion of citable references in your catalog increase by 200% or 1,000% if you include access to articles in books? The answers are critical for planning.

THE PRESENT STUDY

Using samples, we try 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.

STUDY LIBRARIES

We limited our study to university libraries, which may have general or specialized collections, categories that are not mutually exclusive. Would there be a difference in the volume of analytic catalogs in the two types of libraries? Roskilde University Library in Denmark, covering sciences, technology, social sciences, and the humanities, is not a specialized library but a general library. Roskilde University Library was founded in 1972 and has approximately 450,000 volumes. On the other side we have an extremely specialized library, the Special Collection at the National Library of Education in Denmark, covering exclusively education and psychology. This collection was founded in 1887 and has approximately 350,000 volumes that are split into two collections: those published before and after 1990.

SAMPLING TECHNIQUE

Two sampling techniques (one for each library) were required to achieve random samples. In the National Library of Education we selected using physical place-

ment on the shelves ("book number four from the left on each shelf"). Only the actual volume selected was examined, even if it was part of a multivolume work.

To sample from the three subject libraries at Roskilde University Library, we used machine-generated numbers of bibliographic records ending at random digits (e.g., records ending with the digits "82"). In this case, if records represented a multivolume work all volumes were inspected.

The two sampling techniques yield the same total number of articles or citable works in the catalog. The numbers of composite works and of articles within composite works vary depending on whether we count volumes or records; however, as the multivolume records constitute only 1% to 2% of the records in the analyzed library catalogs, we make no distinction here between the results from the two sampling methods because the resulting errors are smaller than the 95% confidence interval (see table 1).

It is assumed that none of these selection rules correlate with the parameters selected for this investigation: the proportion of composite works and the number of articles in these books. Books not found on the shelves were reserved for the study for four months. The period of loan was one month. If they were not received after four months, we excluded them.

To count the composite works, we counted only books containing tables of contents with at least two separately titled works by the same or different authors. To count the works, we counted separately titled and authored works in the tables of contents. No restrictions were made on the number or the length of works counted. Introductions, prefaces, and other generic articles were excluded.

DATE OF PUBLICATION

If either publication or acquisition praxis have changed over time, this could influence the proportion of composite works in the library and the average number of articles in these books. Therefore the books from the National Library of Education were divided into two categories:

books published before and after 1990, to learn whether there is any difference between the proportions of composite works and articles.

LANGUAGE

Because the publication language for international congresses is predominantly English, books in English were examined separately in the National Library of Education to see whether the number of composite works or the number of articles has been influenced by the publication language.

SUBJECT FIELD

Previous studies indicate significant variations between different subject fields with respect to the volume of composite works and the number of articles in these composite works (Weintraub and Shimoguchi 1993, 176). But their data analysis indicates also that the sample for investigation, though selected randomly, might not be typical for the total collection. They consequently ask for further studies of the influence of subject field. We compared science, social sciences, and the humanities with respect to the volume of composite works and the number of articles in the composite works. To do this we looked at a comparison of the holdings of science and social sciences books at Roskilde University Library and the essentially humanistic holdings of books at the National Library of Education.

RESULTS

The results are presented in table 1, with the results from the Santa Ana College Library study (Hoffman and Magner 1985) added in the first row and the results from the San Diego State University Library study (Weintraub and Shimoguchi 1993) added in the second row. The statistical deviation measures are the conventional 95% confidence interval.

The studies of the two Danish university libraries with respect to date of publication, publication language, specialized versus general libraries, and subject field

indicate no or very weak dependence on these parameters. Only the difference between the mean number of articles in composite works in the sciences and the social sciences at Roskilde University Library proves to be significant within the 95% confidence interval.

DISCUSSION AND CONCLUSION

The results from the two American studies and the present study are expected to reflect not only the influences under consideration from subject fields, publication language, publication date, and specialized versus general libraries, but also the influence of acquisition policies at the different libraries. In light of the immense differences between the four libraries, the relatively constant character of the results is striking. The proportion of composite works is between 10% and 20%. The number of articles in the composite works varies from 20 to 30 articles per book—highest for the sciences and the English-language publications and lowest for the social sciences. This implies that the libraries under consideration can add access to between 200% and 600% more works to their catalog without buying one book more, just by adding the tables of contents of their composite works. This is a challenge and an immense increase in the number of access points in the catalog.

But the improved description of the composite works—because of the large amount of text in the individual bibliographic records—may introduce low precision and information overload if analytics are simply added to the catalog. Therefore, most libraries could profit by handling this improvement in separate files to be accessed by the choice of the user.

WORKS CITED

- Atherton, P., and others. 1978. *Books are for use. Final report of the Subject Access Project to the Council on Library Resources.* Subject Access Project, School of Information Studies, Syracuse University, New York. Report IST-10.

The State
of
the Art



...working
smart

*Information Industry
leaders in providing
subscription services,
article delivery & library
automation software.*

- REMO
- ROSS
- Renewal Express
- Financial Planner
- UnCover
- BACKSERV

READMORE ACADEMIC SERVICES

700 Black Horse Pike, Ste. 207
Blackwood, NJ 08012
Phone: 1-800-645-6595
Fax: 609-227-8322

- Beatty, S. 1991. ESP at ADFA after five years. *Cataloguing Australia* 17, no. 3/4: 65-92.
- Byrne, A., and M. Micco. 1988. Improving OPAC subject access: The ADFA experiment. *College & research libraries* 49: 432-41.
- Cochrane, P. Atherton. 1985. *Redesign of catalogs and indexes for improved online subject access. Selected papers of Pauline A. Cochrane*. Phoenix: Oryx.
- Dillon, M., and P. Wenzel. 1990. Retrieval effectiveness of enhanced bibliographic records. *Library hi tech* 31, no. 3: 43-46.
- Hoffman, H. H., and J. L. Magner. 1985. Future outlook: Better retrieval through analytic catalogs. *The journal of academic librarianship* 11: 151-53.
- Knutson, G. 1991. Subject enhancement: Report on an experiment. *College & research libraries* 52: 65-79.
- Lancaster, F. W., and others. 1991. Identifying barriers to effective subject access in library catalogs. *Library resources & technical services* 34: 377-91.
- Lancaster, F. W., C. Elliker, and T. H. Connell. 1989. Subject analysis. *Annual review of information science and technology* 24: 35-84.
- Larson, R. R. 1991. The decline of subject searching: Long-term trends and patterns of index use in an online catalog. *Journal of the American Society for Information Science* 42: 197-215.
- Michalak, T. J. 1990. An experiment in enhancing catalog records at Carnegie Mellon University. *Library hi tech* 8, no. 3: 33-41.
- Poulsen, C. 1990. Subject access to new subjects, specific paradigms and surveys: PARADOKS-registration. *LIBRI* 40: 179-202.
- Van Orden, R. 1990. Content-enriched access to electronic information: Summaries of selected research. *Library hi tech* 8, no. 3: 27-32.
- Weintraub, T. S., and W. Shimoguchi. 1993. Catalog record contents enhancement. *Library resources & technical services* 37: 167-80.