

AD-A162 202

USING ONLINE DATABASES TO DETERMINE THE CORRELATION
BETWEEN RANKED LISTS OF JOURNALS(U) DEFENSE
COMMUNICATIONS AGENCY WASHINGTON DC TECHNICAL AND MAN
D A GUERRIERO 30 DEC 84

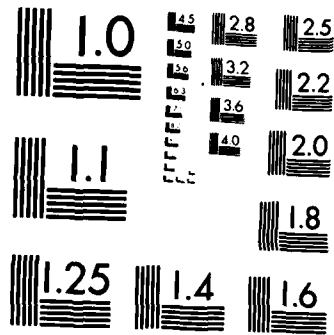
1/1

UNCLASSIFIED

F/G 9/2

NL

END
1/1
1/1



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

2

AD-A162 202

USING ONLINE DATABASES TO DETERMINE
THE CORRELATION BETWEEN RANKED
LISTS OF JOURNALS

by

Donald A. Guerriero

Defense Communications Agency
Personnel and Administration Directorate
Employee Development Division
Technical and Management Information Center

Washington, DC 20305

DECEMBER 30, 1984

DTIC
ELECTE
DEC 04 1985
S D
D

DTIC FILE COPY

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

TABLE OF CONTENTS

Abstract.....1
 Introduction.....2
 Background and Significance.....5
 Procedures.....9
 Results.....11
 Discussion, Implications and
 Recommendations.....18
 Bibliography.....20

Accession For	
NTIS CRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and/or Special
A-1	

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT UNLIMITED	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Defense Communications Agency	6b. OFFICE SYMBOL (If applicable) H396	7a. NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State and ZIP Code) Defense Communications Agency Technical and Management Information Center Washington, D.C. 20305		7b. ADDRESS (City, State and ZIP Code)	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Defense Communications Agency	8b. OFFICE SYMBOL (If applicable) H396	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c. ADDRESS (City, State and ZIP Code) Defense Communications Agency Technical and Management Information Center Washington, D.C. 20305		10. SOURCE OF FUNDING NOS.	
11. TITLE (Include Security Classification) SEE REVERSE		PROGRAM ELEMENT NO.	TASK NO.
		PROJECT NO.	WORK UNIT NO.
12. PERSONAL AUTHOR(S) GUERRIERO, Donald A.			
13a. TYPE OF REPORT Final	13b. TIME COVERED FROM 840701 TO 841230	14. DATE OF REPORT (Yr., Mo., Day) 841230	15. PAGE COUNT 23
16. SUPPLEMENTARY NOTATION			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB. GR.	
5	2	SEE REVERSE	
19. ABSTRACT (Continue on reverse if necessary and identify by block number) SEE REVERSE			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> DTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED	
22a. NAME OF RESPONSIBLE INDIVIDUAL GUERRIERO, Donald A.		22b. TELEPHONE NUMBER (Include Area Code) (703) 692-0373	22c. OFFICE SYMBOL H396

- 11; TITLE: Using Online Databases to Determine the Correlation Between Ranked Lists of Journals
- 18: SUBJECT TERMS: *Keywords*
Information Science, Information Theory, Information Retrieval, Libraries, Library Automation, Serials, Serials Control, Citation Analysis, Periodicals
- 19: ABSTRACT:

This research showed the feasibility of using online databases to help librarians in the selection of journals for the Defense Communications Agency. The purpose of the study was to use citation analysis and statistical testing in journal selection. Bibliographic databases, available to most technical libraries, were used in the research. Development of a selection policy for journals requires authoritative evidence from several sources to justify the cost of the journals selected. The research procedures used in this study included the compilation of a list of 157 technical journals used in the Defense Communications Agency Technical and Management Information Center and the searching of each of two science and technology databases with two predetermined strategies. The databases were searched using, first, a subject search limited to 157 journals in the technical library and, secondly, the same subject search without the journal limitation. The search results were analyzed to determine citations per journal and were rank-ordered by number of citations. The correlation between the two rank-ordered lists on each database was determined. A high degree of correlation was found between the citations produced by the search limited to the Technical and Management Information Center holdings and the citations produced without this limitation. The results of the research implied that the procedure may be useful in developing journal selection policies for the Defense Communications Agency. Further research is required to determine if the same technique with variations in the subjects searched and databases used would produce similar findings. Additionally, this method should be combined with other journal selection techniques for optimum results.

ABSTRACT

This research showed the feasibility of using online databases to help librarians in the selection of journals for the Defense Communications Agency. The purpose of the study was to use citation analysis and statistical testing in journal selection. Bibliographic databases, available to most technical libraries, were used in the research. Development of a selection policy for journals requires authoritative evidence from several sources to justify the cost of the journals selected. The research procedures used in this study included the compilation of a list of 157 technical journals used in the Defense Communications Agency Technical and Management Information Center and the searching of each of two science and technology databases with two predetermined strategies. The databases were searched using, first, a subject search limited to 157 journals in the technical library and, secondly, the same subject search without the journal limitation. The search results were analyzed to determine citations per journal and were rank-ordered by number of citations. The correlation between the two rank-ordered lists on each database was determined. A high degree of correlation was found between the citations produced by the search limited to the Technical and Management Information Center holdings and the citations produced without this limitation. The results of the research implied that the procedure may be useful in developing journal selection policies for the Defense Communications Agency. Further research is required to determine if the same technique with variations in the subjects searched and databases used would produce similar findings. Additionally, this method should be combined with other journal selection techniques for optimum results.

Introduction

The Defense Communications Agency Technical and Management Information Center subscribed to over 500 serials. The Library Director believed that many of these were of poor quality, of questionable importance in the scientific and technical fields the Agency is involved in, and that most were rarely used. The Technical and Management Information Center did not have a selection policy and, to the knowledge of the Director, no scientifically valid study was ever conducted to determine if the current selections were appropriate for the Agency.

The purpose of this research was to begin the process of developing a journal selection policy to meet the needs of the Defense Communications Agency for the next five years. This research to determine if online databases, citation analysis and statistical testing would yield significant results to help in the selection of journals for the Defense Communications Agency.

The Library Director believed that earlier selection policy, or lack thereof, resulted in a collection of scientific and technical journals that had no relationship to (1) the relative importance of journals in the subject areas of concern to the Agency, the quality of the journals, nor to the current needs of the Agency technical staff. There was an immediate need to

develop a selection and de-acquisition policy for technical and scientific journals in the Technical and Management Information Center. One method to do this was to use the techniques of citation analysis of secondary literature.

Citation analysis of secondary literature could be done by the production of lists of journals ranked by the number of citations in secondary literature. By specifying a particular subject, searching for it in a secondary source that indexes and abstracts articles in the subject field, counting all citations discovered, and ranking the journals by their frequency of citation, it would be possible to develop a ranked list of journals in a specific subject. This list would show the relative position of all journals indexed and abstracted by the secondary source. The presumption is that, if the secondary service is carefully selected, the rank-order list will provide one measurement of the relative importance of specific journals in a subject field.

A problem arises, however, in that this list would rank all journals indexed and abstracted by the secondary source if an article were cited. While this list would be useful, there was a

need to compare the list with one generated using only the journals available in the Technical and Management Information Center if the data was to be useful in developing a selection policy.

Using manual methods to develop the rank-order lists and to correlate them would be a time-consuming and difficult, if not impossible, task for a researcher. If, however, computerized databases and other computer tools available, could be used, this problem might be overcome.

This study was a first step in a larger research project that will involve citation analysis of primary literature and journal use pattern studies. While not part of this study, this additional research will eventually be necessary to develop guidelines for the selection and de-acquisition of journals in the Technical and Management Information Center.

There were two research questions to be answered in this study. (1) Is it feasible to use online databases to produce rank-ordered lists of journals in particular subject areas? (2) What is the correlation between: (a) a rank-ordered list of journals, in a specific subject area, that have been indexed and abstracted by leading indexing and abstracting services with readily available online databases, and (b) a rank-ordered list of journals currently subscribed to by the Technical and Management Information Center, using the same subject area criteria?

Background and Significance of the Study

The Defense Communications Agency (DCA) does systems engineering for the Defense Communications System (DCS) to insure that the DCS is planned, improved, operated, maintained, and managed effectively, efficiently, and economically to meet the long-haul, point-to-point, and switched network telecommunications requirements of the National Command Authorities, the Department of Defense (DoD), and other Federal agencies. Secondly, DCA provides system engineering and technical support to the National Military Command System (NMCS), Minimum Essential Emergency Communications Network (MEECN), and the Worldwide Information Network (WIN). Finally, the Agency does system architecture functions for Military Satellite Communications and ADP support to the Joint Chiefs of Staff, Secretary of Defense, and other DoD components.

Within DCA the Technical and Management Information Center supports more than 3500 Washington, D.C. area DCA employees with information from a variety of sources, including scientific and technical periodicals and other serial publications. The Technical and Management Information Center identified about 500 serials being received, among which there were about 160 that might be best called scientific and technical journals or

technical magazines. There was a need to evaluate this collection to determine if it met the needs of the technical staff of the Agency.

In late 1983, the Technical and Management Information Center consolidated the holdings of five libraries into two Technical and Management Information Centers, took over all serials control functions, including the selection, or approval of selection, of all journals purchased for the library system. Additionally, management ordered the Technical and Management information Center to reduce costs by eliminating duplicate or unneeded subscriptions. This research would be used as part of a larger study to determine journal requirements of the Defense Communications Agency

Attempts to use citation analysis as an objective measure of a journal's importance has been done for years. The first major paper on the method appears in Gross and Gross (1927:385-89). The method was simply to select key publications in a subject areas, record the citations in the publications, and rank the cited documents by their count. The method no longer suffices as the body of literature has expanded and automated methods have become available. Using citation analysis as a selection tool, however, is more recent and more controversial.

According to Bradford (1948:1-156) a small number of journals account for a large percentage of the important articles on a particular subject. Coming to be known as "Bradford's Law" this theory resulted in further research to determine if citation analysis could be used in journal selection and collection management. Brooks (1971:458-461) developed a formula to based on Bradford's Law for calculating the "optimum P% library" of journals. Garfield (1972:471-479) explained how citations of papers in original sources could be used to evaluate journals and later showed how citation analysis could be used for selection purposes (1972:5-6). Scales (1976:17-25), found that citation analysis was not a reliable indicator for journal selection. Pan (1978:29-35) found that there was a positive correlation between journal citation and journal usage in libraries and felt that the Scales study did not contradict these finding because the studies were conducted in two different types of libraries. Scales' study was conducted at the National Lending Library that filled requests that other libraries could not. Scales study reflected only filled requests of the patrons of the libraries participating in the study. Line (1978:313-317) flatly stated that citation analysis for the practicing librarian trying to select journals was totally useless. Line did, however, suggest that this statement was a hypothesis that could easily be tested. Boyce and Pollens (1982:29-36) found that there was

little correlation between a Bradford-ranked list of journals in mathematics and a list of the same journals ranked by quality.

Despite conflicting studies on the usefulness of citation analysis, there is no doubt that citation analysis using secondary, indexing and abstracting services will provide a rank-order list of the most often cited journals in a particular field. A similar rank-order list of journals in a library should have positive correlation with the first list for at least the core publications since it is assumed that any librarian would at least want to be selecting the top producing journals. Selection of other journals that might not be indexed and abstracted, such as local publications, newsletters, new and high-quality publications can and should be done by other means.

Procedures

Research Hypothesis

There were two research hypotheses. The first was that rank-ordered lists of journals could be developed using online databases, word processing equipment, and UNIX. The second research hypothesis was that no correlation would be found between rank-ordered lists of all journals included in a database and rank-ordered lists of journals available in the Information Center when such lists are developed for a specific subject area.

Definition of Terms

Journal - a publication appearing periodically such as a technical magazine or scholarly publication. Books, newsletters, government documents, conference proceedings do not fall within this definition.

Technical and Management Information Center - in effect the library of the Defense Communications Agency. The terms Technical and Management Information Center, Information Center and Library are often used interchangeably in the Agency.

Citation analysis - a method of documentary research based on the assumption that the act of citing an author is significant.

Online databases - in the present study these are limited to computerized bibliographic databases that are accessed through a major vendor such as Dialog Information Services.

Selection - the process of determining what journals will be acquired for the library.

Rank-ordered lists - lists of journals in order of frequency of citation.

UNIX - a computer operating system and in this study various programs that are available to process data.

Technical staff - in the Defense Communications Agency the technical staff is largely made up of computer scientists, electronic or electrical engineers, computer programmers or analysts, communication specialists, and operations researchers.

Limitations of the Study

The primary limitation of the study was that the comparison of the rank-ordered lists would only apply to the data from the organization under study. Additionally, only one subject search was performed. The research would only show whether this methodology is feasible and whether it can and should be repeated with other subject searches. If the methodology used showed promise, it might be applicable to other libraries that have access to similar computer resources. In this study there was also a limitation on resources. Online database are costly to use and thus do not lend themselves to indiscriminate use for research purposes.

Basic Assumptions

The basic assumption was that there are online indexing and abstracting services that cover the fields of interest and that the journals available in the library would be indexed and abstracted by those services. The research also assumed that the study could be limited to the time and resources available. Also assumed was that a single subject search would yield information to test the hypotheses.

A list of all periodicals in the Technical and Management Information Center was compiled. Journals and technical magazines considered to be scientific or technical in orientation were isolated and a separate list of those items was made. The manuals of the major vendors of online services available to the Technical and Management Information Center (Lockheed Dialog, SDC Orbit, and BRS) were consulted to determine the databases to use for this research. Two databases on SDC Orbit, Compendex and Inspec, indexing and abstracting articles in the areas of interest to the Defense Communications Agency technical staff, were selected. Each of these databases, using the SDC software, permitted the researcher to perform a subject search and to search by title of the periodical, using *Codens* in place of full titles. Additionally, the SDC software permitted the researcher to store search strategies of the length required for this study. Finally, the SDC version of the Compendex and Inspec database simplified offline printing and sorting by journal title.

The databases were searched with a subject search. In this study the term "Local Area Network" was used, although any appropriate subject could have been chosen. The final search strategy limited the search to journals and publication years 1980 to 1984.

Two searches were conducted on each database using (1) a strategy that limited the search to journals in the library and (2) a strategy that covered the entire database. An offline citation list was generated by SDC Orbit for each search on each database. The citation list, was sorted by source and included only the title of the article cited and the source.

Four printed citation lists resulted from this procedure. Each list was analyzed and all sources that did not produce at least two citations were eliminated. The resulting four lists were separately rank-ordered by the number of citations produced for each journal. Statistical analysis of the rank-ordered lists was then performed.

The null hypothesis was tested that there was zero correlation in the rankings of the journals that produced more than one citation in the citation list using the search strategy limiting retrieval to journals in the library and the citation list that resulted from the search conducted without this limitation.

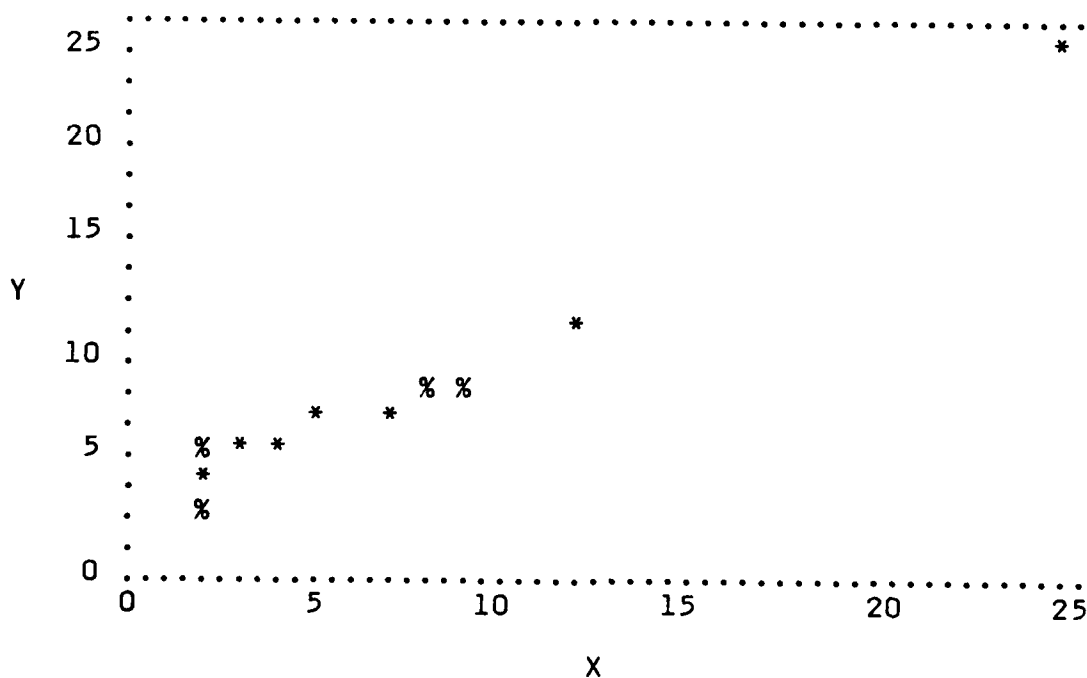
The parametric correlation technique, the Pearson-Product-Moment correlation, was used to determine the coefficient of correlation between the variables. The correlation was tested for significance at the .01 level.

Results

Searching the database Compendex on SDC Orbit using the subject "Local Area Networks," limiting the search to 157 journals in the library and to publication years 1980-1984 resulted in a list of 17 journals with more than one citation per journal. This variable was called "DCA Periodicals" and labeled "x". A second search of the Compendex database using the subject "Local Area Networks" and limiting the publication years to 1980-1984, but not limiting the search to journals in the library resulted in a citation list of 36 periodicals with more than one citation. This variable was called "All Database" and labeled "y". The variables x and y were rank-ordered and the number of citations retrieved in the 17 rank-ordered journals of variable x were correlated with the number of citations retrieved in the top 17 rank-ordered journals of variable y. The table below illustrates this procedure.

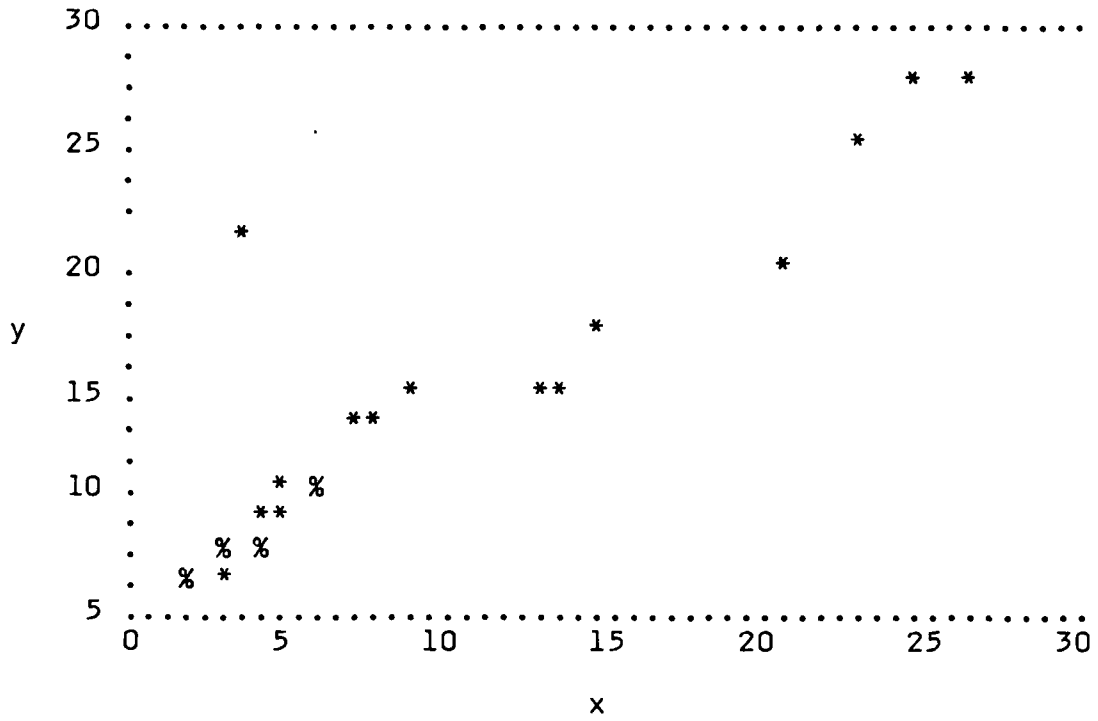
Rank	DCA PERIODICALS (No. Citations) x	ALL DATABASE (No. Citations) y
1	25	25
2	12	12
3	09	09
4	09	09
5	08	08
6	08	08
7	07	07
8	05	07
9	04	06
10	03	06
11	02	06
12	02	05
13	02	04
14	02	03
15	02	03
16	02	03
17	02	03

The correlation of coefficient between x and y was 0.979219, a high and extremely dependable relationship, significant at the .01 level. The correlation was graphically shown by producing the following scatter plot, using the S statistical package on UNIX.



Similar results were obtained using the same procedures on the database Inspec. The correlation was 0.987354, an equally high and extremely dependable relationship significant at the .01 level.

A scatter plot of the Inspec correlation is shown below.



Discussion, Implications and Recommendations

The study showed that it is feasible and practical to produce rank-ordered citation lists using online databases readily available to technical librarians. The procedure used to limit the database search to specific journals proved to be somewhat difficult, however. Offline searches using microcomputers should be investigated as an alternative to the procedure used in this study. The citations lists were easily produced once the search strategy was developed and rank-ordering was also easily done. The statistical analysis was done with the S computer statistical package on UNIX, but could be done by another statistical package or by manual methods.

This study showed a high correlation between the number of citations produced by a search of the databases limited to the 157 library periodicals and the search without this limitation. This result appears to show that the library holds technical journals necessary to produce enough citations when the online search is limited to periodicals held by the library. Subscribing to additional journals that produced more than one citation, that are not held by the library, would not raise the number of citations by a significant amount.

Further research is necessary to form an opinion on the validity of the technique using different subject searches and the same databases used in this study and perhaps on other databases. Additionally, further research should be conducted using other statistical tests and use surveys are also required.

BIBLIOGRAPHY

- Boyce, Bert R.; Pollens, Janet Sue, "Citation-Based Impact Measures and the Bradfordian Selection Criteria," *Collection Management*, 4 (3) Fall 1982, pp. 29-36.
- Brooks, B.C., "Bradford's Law and the Bibliography of Science," *Nature*, Vol 224, December 6, 1969 pp. 653-656.
- Bruer, John T.; Goffman, William; Warren, Kenneth, "Selective Medical Libraries and Library Networks for Developing Countries," *American Journal of Tropical Medicine and Hygiene*, 30 (5) September 1981, pp. 1133-1140.
- Dhawan, S.M.; Phull, S.K.; Jain, S.P., "Selection of Scientific Journals: A Model," *Journal of Documentation*, 36 (1) March 1980, pp. 24-32.
- Engineering Information, Inc. Publications Indexed for Engineering. New York: Engineering Information Inc., 1982.
- Farber, Evan Ira, "Limiting College Library Growth: Bane or Boon?" *Journal of Academic Librarianship*, 1 (5) November 1975, pp. 12-15.
- Gordon, Michael D., "Citation Ranking Versus Subjective Evaluation in the Social Sciences," *Journal of the American Society for Information Science*, 33 (1) January 1982, pp. 55-57.
- King, Donald; McDonald, Dennis D.; and Roderer, Nancy. *Scientific Journals in the United States*. Stroudsburg, Pennsylvania: Hutchinson Ross Publishing Company, 1981.
- Institution of Electrical Engineers. *List of Journals and other Serial Sources*. Surrey, G.B.(?): Institution of Electrical Engineers, 11983
- Koenig, Michael E.D., "On-Line Serials Collection Analysis," *Journal of the American Society of Information Science*, 30 (3) May 1979, pp. 148-153.
- Kriz, Harry M. "Subscriptions vs. Books in a Constant Dollar Budget," *College and Research Libraries*, 39 (2) March 1978, pp. 105-109.

- Line, Maurice B. "Rank Lists Based on Citation and Library Uses as Indicators of Journal Usage in Individual Libraries," *Collection Management*, 2 (4) Winter 1978, pp. 313-316.
- Pan, Elizabeth, "Journal Citation as a Predictor of Journal Usage in Libraries," *Collection Management*, 2 (1) Spring 1978, pp. 29-38.
- Peters, Andrew, "Evaluating Periodicals," *College and Research Libraries*, 43 (2) March 1982, 149-151.
- Sandison, A., "Documentation Note: References/citations in the Study of Knowledge," *Journal of Documentation*, 31 (3) September, 1975, pp. 195-198.
- Scales, Pauline A., "Citation Analysis as Indicators of the Use of Serials: A Comparison of Ranked Title Lists Produced by Citation Counting and from Use Data," *Journal of Documentation*, 32 (1) March 1976, pp. 17-25.
- Singleton, Alan, "Journal Ranking and Selection: A Review in Physics," *Journal of Documentation*, 32 (4) December, 1976, pp. 258-289.
- Windsor, D.A., "Rational Selection of Primary Journals for a Biomedical Research Library: The Use of Secondary Journal Citations," *Special Libraries.*, 64 (10) October 1973, pp. 446-451.

END

FILMED

1-86

DTIC