The scope of this bibliography has been limited to literature dealing with the design, testing and evaluation of information storage and retrieval systems. A number of papers describing the implementation of specific systems were selectively included because they contained substantial or innovative material on evaluation; however, the bibliography generally emphasizes techniques applicable to a wide variety of systems. Similarly, articles dealing with specific products and services have been included when they showed broad implications for evaluation. On the other hand, discussions of specific library systems have been excluded as they were considered outside the scope of the bibliography. You will not, therefore, find such articles as those dealing with "objective tests of library performance." The last comprehensive bibliography in the field was compiled by Madeline Henderson and published in 1967. For that reason, the earliest citations in this compilation are for 1967. (Author/SJ)
EVALUATION OF INFORMATION SYSTEMS:
A BIBLIOGRAPHY, 1967-1972

Beth I. Krevitt
Government Services Division
Aspen Systems Corporation

and

Belver C. Griffith
Graduate School of Library Science
Drexel University

Published in cooperation with the

ERIC Clearinghouse on Library and Information Sciences
1140 Connecticut Avenue, N.W., Suite 804
Washington, D.C. 20036

This publication was prepared pursuant to a contract with the National Institute of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their judgment in professional and technical matters. Points of view or opinions do not, therefore, necessarily represent official National Institute of Education position or policy.
INTRODUCTION

Design

The scope of this bibliography has been limited to literature dealing with the design, testing and evaluation of information storage and retrieval systems. A number of papers describing the implementation of specific systems were selectively included because they contained substantial or innovative material on evaluation; however, the bibliography generally emphasizes techniques applicable to a wide variety of systems. Similarly, articles dealing with specific products and services have been included when they showed broad implications for evaluation. On the other hand, discussions of specific library systems have been excluded as they were considered outside the scope of the bibliography. You will not, therefore, find such articles as those dealing with "objective tests of library performance."

The last comprehensive bibliography in the field was compiled by Madeline Henderson and published in 1967. For that reason, the earliest citations in our compilation are for 1967, although no systematic attempt was made to remove those citations duplicating Henderson's work. The year 1972 was considered an expedient limiting date, thereby permitting the bibliography to be current and yet remain a reasonable size. Final preparation of the bibliography began in early January 1973, thus excluding publications from late 1972. The authors felt that this cutoff was necessary if the bibliography was to be printed in 1973.

Methodology

The basic tools used in compiling this material were: Research in Education (via ERIC/CLIS on-demand searches and monthly accession lists), Library Literature, and Information Abstracts. Liberal use was made of bibliographies in the Annual Review of Information Science and Technology (Carlos A. Cuadra, ed.) All material (monographs, technical reports, journal articles, rejoinders, etc.) have been included regardless of possible duplication of content in the hope that this degree of noisiness in our approach will increase the potential user's ability of actually obtaining a copy of the material desired. Occasionally both PB and ED numbers have been available; when this has occurred both have been included. Also, if an author has flatly stated that two publications are identical, this has been noted.

Each citation is listed under one (and only one) category within the classification schedule. (See page 5 and 6 for the categories used in the classification scheme.) Although this limits the access to a citation to only one point, the authors felt that a multiple access point bibliography with such a limited scope would become highly repetitious. The hardest material to classify, of course, was the published monographs. These, therefore, have been grouped together outside the classification scheme.

Some General Impressions

We started this bibliography in the hope that we would gain a deeper understanding of the problems of evaluating information systems. At the beginning, we had expected to read, and in the words of Licklider, 'interact deeply,' with this literature. It proved far too large, and our industry and the time available did not permit us that luxury. We are a little disappointed at being unable to carry through our original plan.

We are also somewhat disappointed in the literature that we have been inspecting over these months. It is piecemeal, noncumulative, and in a rather fundamental sense, moribund. It is a reflection upon our own field that there is not more integration of work, greater breadth of perspective in this area, and better scholarly analysis and review of the work. The stagnant quality of the literature has another origin, namely, the apparent lack of any recent and substantial funding to do the massive studies required to make basic further contributions to this field. Therefore, there is a growing literature that consists largely of reanalyses of, for example, the Cranfield data of the application of indexing to retrieval.

The literature which this bibliography contains seems to have advanced little beyond the work of Cleverdon, Guiliano, Salton and Swets in the area of retrieval; little beyond the work of Bradford, Brookes, Fairt:lorne and Price in the structure of information; and little beyond the work of Orr, Allen, Kent, Wolek, Garvey and Griffith, in the analysis of the benefits of information to the user.

The most distinguished effort is the work of the group of researchers and information specialists that clustered in Westat between 1968 and 1970, meaning, of course, Lancaster, King, and Bryant, and in the new focus of their work on cost-effectiveness. The single most influential article, the 1971 JASIS article by Lancaster,2 was hailed by the community with a sigh of relief. The information science community has long received beatings on the simple straight-forward basis that they could not demonstrate much substantial

---

benefits to users, and Lancaster offered them an alternative which can be summarized, perhaps a little negatively, as "we may not be doing anything of ultimate value (and it may not be up to us to discuss 'ultimates') but what we are doing we are doing very well."

The results are not in on whether this new approach will substantially affect the relation of the information science community to its funders. (We are not optimistic.) It has clearly created great benefits within the community in terms of calling attention to the factors common to most document systems and by emphasizing the intrinsically probabilistic nature of information and its handling. The only effect of the cost-effectiveness approach that we deplore is that it effectively separates an "evaluation" literature from a "user" literature, to the possible detriment of both. The concept of the cost-effectiveness of information systems appeared just at the point that it was becoming very evident 1) that most information systems created for scientists took little account of either the informal or conceptual structure of science, which clearly guides the scientist's search for information, 2) that only the most preliminary data on the information habits and needs of workers in applied science and technology had been gathered. In general, the "user" research had begun to raise questions about the "document" and "retrieval" orientations of precisely the type of systems the Westat group were best prepared to evaluate.

Cleverdon wrote in a recent review (Journal of Documentation of the major monograph of the Westat school of evaluation.3 "It is clear that the authors thought deeply about the subject without doing very much practical work...." However, this group has done a great deal of practical work, or what passes for it to government funders, on evaluations of real systems. More in line with our own thinking, Swanson wrote (in Library Quarterly) regarding the same monograph, "in a sense it is a handbook, almost a 'cookbook', of what to measure and how to utilize such measurements (to describe) the performance of the system. Some attention is also given to cost analysis, with detailed examples." The implicit contrast between these comments arises naturally from the Westat approach. On one hand, the approach does not really help the practitioner in his rough and ready dialogue, filled with all sorts of unspoken assumptions and biases, between himself and his governing body or agency. This interaction, which the author (Griffith) has for pedagogical reasons called the "evaluation dialogue" almost never rises above the restrictions of "local ground rules," even for the largest information centers and services.

The other, and much more sweeping criticism, is best understood at a very general level: Information systems rarely offer a single service or product and there are, presently, no formal ways of equating and optimizing total output. Similarly, the user is rarely satisfied with a single product or service, and there are, presently, no formal ways of equating and optimizing total desired input to the user. With this as a given, the information practitioner insures, basically, that the few facts he gathers and the judgments he makes are the facts and judgments welcomed by the governing group. Again, if you accept the assumptions regarding information output (for the system) and information input (to the user), the use of elaborate techniques, however detailed, will nearly always be inconclusive.

Perhaps the real force behind the development of this literature is that we have no more than the old practitioner-governing body dialogue, only this time the practitioner heads a large system and the governing body is the Federal government. With that level of resources, the facts required and judgments rendered may be very complex indeed, even though the outcome may still be inconclusive.

Finally the reader, if he wishes to get an understanding of the state-of-the-art and professional climate of the present, should read the works of King, Bryant, and Lancaster with care. They are certainly the literature to which future work will react.
I. Systems analyses of information systems
   A. General evaluation techniques
   B. General design techniques
   C. Modelling and simulation techniques
   D. System performance functions
   E. Cost-effectiveness techniques
      (The economics of information retrieval)

II. Information system components
   A. Evaluation measures
      1) Relevance and recall
      2) Other retrieval effectiveness measures
      3) Applications of statistics to other aspects of systems
   B. File structure and design
   C. Search strategies
   D. Indexing languages and systems
      1) Conceptualization and methodology
      2) Performance and effectiveness
      3) Evaluation
      4) Language devices
      5) Indexer consistency
III. Information products and services

A. User/system interface

B. Current awareness and selective dissemination of information (SDI) systems

C. Real-time, interactive systems

D. Evaluation of other specific information products and/or services
PUBLISHED BOOKS

The following published books are either wholly devoted or largely related to the evaluating of information systems:


JOURNAL ARTICLES AND TECHNICAL REPORTS

I. SYSTEMS ANALYSES OF INFORMATION SYSTEMS

A. General Evaluation Techniques.


Comparative Systems Laboratory. An Inquiry into Testing of Information Retrieval Systems. 3 vols. Edited by Tefko Saracevic and others. Case Western Reserve University, Cleveland, 1968, (PB 179290 or ED023421, PB 180952 or ED 027042, PB 180952 or ED 027041).


Good, I. J. "The decision theory approach to the evaluation of information retrieval systems." Information Storage and Retrieval, 3:2 (April 1967) 31-34.


Saracevic, Tefko. "Selected results from an inquiry into testing of


Vickery, B. C. "In the last analysis all evaluation is subjective." Methodology in Research. Aslib Proceedings 22 (1970) 597-606.


B. General Design Techniques.


Center for Information Services, Phase II: Detailed System Design and Planning. 7 Parts. Institute of Library Research, Los Angeles, California, 1971. (ED 057 496 through ED 057 812).


C. Modelling and Simulation Techniques.


Findlay, Donald C. "Application of the CIPP evaluation model to a center with multiple program areas and levels." Educational Technology 11:10 (October 1971) 43-47.


Levy, Burt; Menden, Werner H. Multiple Test of ABC Methods. Part III,


D. System Performance Functions


E. Cost-effectiveness techniques (the economics of information retrieval)


King, Donald W. Cost/Benefits Analysis Westat Research, Inc., Bethesda, Maryland, 1969. (Study conducted for the Mexican Psychological Association).


Robertson, S. E.; and others. "Standard costing for information retrieval systems: background to a current study." Aslib Proceedings 22 (September 1970) 452-457.

Rolling, L; Pielte, J. "Interaction of economics and automation in a

Rogers, F. B. "Cost of operating an information retrieval service." Drexel Library Quarterly 4 (October 1968) 271-278.


II. INFORMATION SYSTEM COMPONENTS

A. Evaluation Measures
1. Relevance and recall

Barhydt, G.G. "Effectiveness of non-users relevance assessments." Journal of Documentation 23 (June 1967) 146-149.


Cleverdon, Cyril W. Effects of Variation in Relevance Assessments in Comparative Experimental Lists of Index Languages. Cranfield Institute of Technology, 1970.


Cuadra, Carlos A. A Study of Relevance Judgments, 1968, 12 p. (Final draft of talk for 1968 Annual Meeting of the American Psychological Association, San Francisco.)


Davis, M. C.; Linsky, M.D.; Zelkowitz, M. V. "A relevance feedback system employing a dynamically evolving document space." In: Reports on Analysis, Search and Iterative Retrieval, X-1 through X-29. (See under General evaluation techniques - Gerald Salton.)

Dym, E. D. "Relevance predictability. I. Investigation
Rees, Alan M.; Schultz, Douglas J. A Field Experimental Approach to the Study of Relevance Assessments in Relation to Document Searching. Final report to the National Science Foundation. 2 Volumes. Case Western Reserve University, Cleveland, Ohio, October 1967, 305 pp. (PB 176079 or ED 027910; PB 176 080 or ED 027909).


2. Other retrieval effectiveness measures.


Cooper, W. S. "Expected search length: a single measure of retrieval effectiveness based on the weak ordering action of retrieval systems." American Documentation 19 (January 1968) 30-41 (Correction in American Documentation 19 (July 1968) 355.)


3. Application of statistics to other aspects of information systems.


Königórá, M. "Mathematical and statistical methods of noise evaluation in a retrieval system." Information Storage and Retrieval 6 (May 1971) 437-444.


B. File Structure and Design.


Rettenmayer, John W. "File ordering and retrieval cost." Information Storage and Retrieval 8:2 (April 1972) 79.


C. Search strategies.


D. Indexing languages and systems.

1. Conceptualization and methodology.


Jahoda, Gerald; Stursa, M. L. Test of Indexes: A comparison of keyword from title indexes with and without added keywords and a single access point per document alphabetic subject index. Library School, Florida State University, 1969, 61 pp.


Landry, Bertrand Clovis. A Theory of Indexing: Indexing theory as a model for information storage and retrieval, Computer and Information Science Research Center, Ohio State University, Columbus, Ohio, December 1971, 282 pp. (OSU-CISRC-TR-71-13; ED 057843).


2. Performance and effectiveness.


Tell, B. V. "Retrieval efficiency from titles and the cost of indexing." Information Storage and Retrieval 7 (December 1971) 241-243.


3. Evaluation.

"Comparison of indexing systems." Library World 70 (September 1968) 90.


Bloomfield, M. "Evaluation of indexing."

(1) "Introduction" Special Libraries 61:8 (Oct 70) 429-432.
(2) "The simulated machine indexing experiment." Special Libraries 61:9 (Nov 70) 501-507.
(3) "A review of comparative studies of index sets to identical citations." Special Libraries 61:10 (Dec 70) 554-561.


4. **Language devices.**


Swanson, Don R. "Some unexplained aspects of the Cranfield tests of indexing performance factors." *Library Quarterly* 41:3 (July 1971) 223-228.


5. **Index consistency.**

Cooper, W. S. "Is inter indexer consistency a hobgoblin?" *American Documentation* 20:3 (July 1969) 268-278.


III INFORMATION PRODUCTS AND SERVICES

A. User-System Interface

Lancaster, Frederick Wilfrid. "Interaction between requesting and a large mechanized retrieval system." Information Storage and Retrieval 4:2 (June 1968) 239-252.


B. Current-Awareness and selective dissemination of information (SDI) systems

Anderson, R. R.; and others. The Stability of Dynamic Feedback in an
**SDI System.** Ames Institute for Atomic Research, Iowa State University of Science and Technology, Iowa, 1968, 24 pp. (ED 034563).


Davison, P. S. "Operating large-scale, broad coverage CA and SDI services." *The Information Scientist* 6:1 (March 1972) 15.


Jordan, John R. *A Framework for Comparing SDI Systems.* Ames Institute for Atomic Research, Iowa State University of Science and Technology, 1968, 6 pp. (PB 179796; ED 034564; Article by the same name published in *American Documentation* 19:3 (July 1968) 221-272.)


C. Real-time, interactive systems.


Kurtz, Peter; and others. On-line Retrieval II. Informatics, Inc., Bethesda, Maryland, May 1971 (TR71-1202-1; ED 057824).


Timbie, Michele; Coombs, Don H.  An Interactive Information Retrieval System; Case Studies on the Use of DIALOG to Search the ERIC Document File. ERIC Clearinghouse on Educational Media and Technology, Stanford University, California, December 1969, 90 pp. (ED 034431).


D. Evaluation of other specific information products and/or services.


Artandi, Susan; Wolf, E. H.  Effectiveness of Weights and Links in


Lancaster, Frederick Wilfrid. Evaluation of the MEDLAPS Demand


Stevens, Norman D. "MEDLARS: A Summary Review and Evaluation of