

AFOSR 67-2248 HERNER AND COMPANY
2431 K STREET, N. W.
WASHINGTON, D. C. 20037

TELEPHONE
908-3100

CABLE ADDRESS
"HERNCO"

August 9, 1967

Study of Documentation Procedures and Mechanisms

Director of Information Sciences
Air Force Office of Scientific Research
Room 440
1400 Wilson Boulevard
Rosslyn, Virginia 22209

Subject: Final Scientific Report on
Contract AF 49(638)-1424

Gentlemen:

Pursuant to the terms of the subject contract, I am pleased to submit herewith a review of the work performed and a chronological bibliography of the publications produced in whole or in part during the period of the contract.

As noted in progress reports and other communications submitted in connection with the subject contract and predecessor contract, the basis for the work done and reported was wholly pragmatic; this work derived from the results of empirical studies performed over the years to find solutions to problems of general concern in information science and technology and/or of concern to specific clients in Government and industry. In terms of physical products, the project produced six papers, three of which have already been published and three of which will be published within the next several months. The paragraphs which follow are given to a discussion of the derivation, content, and primary implications of these publications.

ASLIB-Cranfield Evaluation Techniques

One of the first study tasks undertaken during the period of the subject contract was an analysis of the major areas of applicability and significance of the evaluation procedures developed by the ASLIB-Cranfield project.¹ This task consisted of a follow-up and re-evaluation of the results of a rather detailed test performed on an index to a portion of the unpublished report collection of the Bureau of Ships Technical Library. In essence, the purpose of this follow-up and re-evaluation was to determine

¹ Herner, S., F. W. Lancaster and W. F. Johanningsmeier, A Case Study in the Application of Cranfield System Evaluation Techniques, Journal of Chemical Documentation, 5, 92-5, 1965.

Distribution of this
document is unlimited

Reproduced by the
CLEARINGHOUSE
for Federal Scientific & Technical
Information Springfield Va 22151

AD 659625

HERNER AND COMPANY

WASHINGTON, D. C. 20037

Director of Information Sciences
Air Force Office of Scientific Research
August 9, 1967

Page 2

just what the technique revealed about the workings and efficacy of an index.

From reading the plethora of reports and papers on the subject, one comes away with the impression that relevance and recall ratios are true figures of merit. However, from our own case study and our review of its significance and results it is clear that relevance and recall ratios, as applied to in vivo indexing and retrieval situations, cannot be taken as either absolute or comparative figures of merit. Properly applied, the technique can furnish an approximation of the amounts of "noise" and "misses" in search products. Carried further, and administered by extremely knowledgeable and astute analysts, the technique can reveal with reasonable accuracy the causes of impurities and/or omissions in these products. However, measured against user responses and preferences, it is difficult to ascribe merit or demerit to high or low relevance or recall. There are circumstances under which the search requestor will sacrifice useful documents in order to avoid having to wade through irrelevant ones; and, conversely, there are circumstances where requestors will accept impurities in search products in order to attain comprehensiveness. Thus, relevance and recall ratios cannot be equated with good and bad in indexes and retrieval systems, as related to use and usefulness.

From our own experience and those of others, probably the most useful product or by-product, of the entire ASLIB-Cranfield exercise is the fact that it prompts the designer and operator to dissect all elements of his system in order to determine why it acts or responds the way it does, and to develop bases for positive change. In this sense, relevance and recall ratios can be used as indices of change, and perhaps this is their greatest significance, although we did find that in applying the postmortem aspect of the ASLIB-Cranfield technique we were forced into a detailed understanding of the human and machine mechanisms involved in the input, throughput, and output aspects of index systems. Thus, inadvertently, the technique becomes a useful tool for systems analysis. A significant portion of our effort, as detailed in the paper cited, was given to developing procedures for converting the technique from an inadvertent to a formal tool of retrieval system analysis.

System Design, Evaluation, and Costing

As an outgrowth of our work with the ASLIB-Cranfield technique, studies of information-gathering and communication patterns, and assignments involving the design, costing, and implementation of information systems and programs,

HERNER AND COMPANY

WASHINGTON, D. C. 20037

Director of Information Sciences
Air Force Office of Scientific Research
August 9, 1967

Page 3

we turned our attention to some of the broader issues of information systems, namely, what systems are, what they do, and how to go about designing and evaluating them in toto. This work resulted in a paper which is shortly to be published in Special Libraries.² Again by means of case studies and in vivo illustration, the paper begins with an attempt to clarify the rather muddled notion of the system concept as applied to information programs and processes, and to establish the relationship and role of the designer and operator. From this baseline, the paper turns to a discussion of means of determining system requirements, as syntheses of organizational policies, purposes, and economics, and the information needs and communication characteristics of the presumed user-group. The paper then proceeds to a discussion of means of extracting and defining system requirements from the foregoing inputs. Interestingly, in the case illustration that we used in discussing means of defining requirements, and in most other illustrations that we could have used, it develops that the lowest priority need is for a retrospective search capability, and the highest priority needs are for (1) means of putting people directly in touch with one another, (2) information about ongoing research projects, and (3) information on current research results. This order of priority, which is becoming increasingly common, is indicative of the fact that activities such as the ASLIB-Cranfield project, dedicated as they are to retrospective retrieval systems, are in fact much ado about not as much as many people think.

From the subject of requirements, the paper turns to the question of the selection of methods or mechanisms for meeting them, which comes down to a matching of alternative means against the jobs to be done, and evaluating them in terms of technical efficiency, probability of user acceptance, and costs. Perhaps the most important element in this matching process, and the one which is most frequently overlooked, is the habits, preferences, and idiosyncrasies of the people who are to be the users of the system. In the approach to design discussed in the paper, we treat the user as an integral part of the system, and take careful account of his probable interaction with it. As a result of some muddled thinking over the years, many designers have been literally scared away from detailed consideration of user interaction, and this has given rise to many expensive white elephants.

The matter of costs has also been the subject of a lot of muddled thinking. An unfortunately large number of people in the field are of the opinion that cost analyses should be based primarily on magnitude of input. However, costs and cost justification are in fact mainly functions of quantities of output. The larger the number of searches that a system performs,

² Herner, S., System Design, Evaluation and Costing—in Plain English, March 1967.

HERNER AND COMPANY

WASHINGTON, D. C. 20037

Director of Information Sciences
Air Force Office of Scientific Research
August 9, 1967

Page 4

the smaller the cost per search; the larger the number of copies of disseminated system outputs, the smaller the cost per copy.

From methods and bases of costing, the paper turns to the question of actual design and design evaluation. Here again, an attempt is made to put the matter of evaluation into context. This is done by showing and discussing the various administrative, fiscal, technical, and user-acceptance levels of evaluation. We found, for instance, in one illustrative design study which we performed on behalf of a client organization that, by ASLIB-Cranfield lights, the organization's existing retrieval system was excellent; however, its cost of operation was so high, and the resistance of its users so strong, that it was totally unrealistic as a viable information tool. Thus, two elements of evaluation, neither concerned with retrieval effectiveness, guided the decision to discard the system.

Review of User Studies

During the period of the contract, considerable time was also given to the bibliographic and interpretative phases of the preparation of a chapter, Information Needs and Uses in Science and Technology, for the second volume of the Annual Review of Information Science and Technology.³ In all, a total of 26 user studies produced during 1965 and 1966 were reviewed. These papers, for the most part, added little to previous knowledge of information-gathering or flow patterns. Two, however, by Allen and by Smith, have important implications for system evaluation and design, training of scientists and engineers in the use of information tools and techniques, and the measurement of cost effectiveness in systems and programs.^{4, 5} Both of these studies deal with empirical analyses of the relationship of information-gathering and communication patterns and levels of work performance among scientists and engineers. Both reveal definite differences in the information tools and techniques used by the high and low performers.

³ Herner, S., and M. Herner, Information Needs and Uses in Science and Technology, Annual Review of Information Science and Technology, v. 2, pp. 1-34, New York, John Wiley & Sons, 1967 (in press).

⁴ Allen, T. J., The Differential Performance of Information Channels in the Transfer of Technology, Cambridge, Mass., Massachusetts Institute of Technology, June 1966.

⁵ Smith, C. G., Organizational Factors in Scientific Performance in an Industrial Research Laboratory, Madison, Wis., Center for Advanced Study in Organization Science, Univ. of Wisconsin, Oct. 1966.

HERNER AND COMPANY

WASHINGTON, D. C. 20037

Director of Information Sciences
Air Force Office of Scientific Research
August 9, 1967

Page 5

A very fundamental point, which is tacitly understood but rarely articulated or pursued, is that the purpose of all information activities and services is to improve or maximize the performance of their clientele. Thus, in measuring effects on performance, we are approaching the ultimate evaluation of a system or program. We are also developing bases for the design of truly meaningful and effective systems and programs with clearly defined goals against which they can be measured. Also, in regard to cost effectiveness, this view of information processes furnishes a basis for determining whether we are really getting our money's worth by telling us what we should be buying to attain the greatest benefits. It also has interesting implications for the effective training of scientists and engineers in the use of information tools, techniques and resources.

We are presently working on a design for a study to define and enlarge upon the techniques of Allen and Smith. We believe that this study, which will involve behavioral scientists, information scientists, survey specialists, and cost accountants, could have a profound impact on future means for determining optimal fiscal allocations for information tools and resources, on system and program design, and on the training of scientists and engineers.

Machine Analysis of Psychiatric Interviews

Another major effort during the period of the contract was given to a study of the applicability of Luhn's autoabstracting autoencoding techniques to the analysis of the contents of psychiatric interviews. This work is being done in collaboration with the Study Center of the Washington School of Psychiatry. Details of progress of the study, which has as its purpose the preparation of statistically-based word distillates of discursive interviews, are given in a paper, Application of Automatic Literature Analysis Techniques to Psychiatric Interviews, which is being refined for submission for publication.⁶

The refinements of the paper are concerned with the evaluation of the validity of the techniques and the kind and degree of assistance they can render the psychiatrist in reinforcing his memory and furnishing hints as to the nature, causes, and stages of pathologies. In order to evaluate the usefulness and significance of test autoabstracting and autoencoding of the first of nine teaching interviews by Wolberg, we are presently distributing protocols to a panel of fifteen psychiatrists and clinical psychologists,

⁶ Herner, S., H. A. Segal, and E. Leyman, Application of Automatic Literature Analysis Techniques to Psychiatric Interviews, Washington, D. C., Herner and Company, 1967.

HERNER AND COMPANY

WASHINGTON, D. C. 20037

Director of Information Sciences
Air Force Office of Scientific Research
August 9, 1967

Page 6

who are being asked a series of questions to determine what they got out of the whole interview that they did not get out of the autoabstracting and autoencoding products and vice versa.⁷ The need for more rigorous evaluation and validation of the techniques became clear from the reviews and criticisms of our paper by several eminent psychiatrists and psychologists to whom it was submitted for comment.

Perhaps more germane to the purposes of the subject contract is a by-product of our work with psychiatric interviews. This by-product has to do with the application of autoencoding techniques to the preparation of Thesauri and SDI profiles. As a result of extensive efforts over the years to develop rational and consistent bases for the design of subject authorities and personal and group profiles, we were naturally led from our work with psychiatric interviews to consider the possible applicability of autoencoding techniques to these areas. Analyzing major Thesaurus-design efforts, such as those of the Engineers Joint Council and Project Lex, we were impressed with their relative subjectivity, unevenness and inconsistency, and with the rigor which would seemingly be injected by a statistically-based technique such as autoencoding. We are presently analyzing and refining techniques to permit us to prepare a Thesaurus based on the statistical analysis of a field of science. What we are doing is a subject of controversy, in that many people in the field claim that the mere fact that words and combinations of words crop up frequently in the literature of a field is not an indication of their significance. However, it has been shown, at least in the pure or basic sciences, that the prime users of a literature and indexes to it are the people who contribute most frequently to it. Thus, we think we are on the right track.

The rationale behind the notion of basing profiles on the writings of individuals and groups is somewhat akin to that for basing the Thesaurus of a field on its literature. To date, there have been myriad schemes, of varying plausibility, for developing group and individual profiles for SDI systems. As in the case of Thesaurus development, these have been largely subjective. Based on the aforementioned writer-reader relationship, it would seem entirely plausible to base profile design on statistical analyses of representative cross sections of the writings of the profilees. This should produce rigorous, consistent, and up-dateable profiles, and we intend to find out if this is true.

Functional Aspects of Information Networks

Another, somewhat smaller study performed during the contract period

⁷ Wolberg, L. P., The Technique of Psychotherapy, pp. 688-780, New York, Grune and Stratton, 1954.

HERNER AND COMPANY

WASHINGTON, D. C. 20037

Director of Information Sciences
Air Force Office of Scientific Research
August 9, 1967

Page 7

was an outgrowth of an earlier study, performed on behalf of the Office of Science and Technology, to develop a recommended design for a national medical library and information system.⁸ The specific focus of the outgrowth of the prior study was the interaction between national networks and individual scientist and small, undermanned libraries. In the typical blue-sky approach to network concepts, there is usually incorporated the tenuous assumption that by setting up or augmenting libraries and information centers and connecting them telephonically or electronically the information problems of professional and scientific practitioners and understaffed or staffless libraries will, as if by magic, be solved.

Focusing on the extensive data and background information gathered in the course of our network design, we considered in detail the probable relationships of networks to individuals and small client-libraries. What emerged from these considerations was an analysis of the pitfalls and indicated precautions and remedies involved in making networks and network schemes realistically responsive to the needs of their individual and institutional clientele. This analysis is summarized in a paper presented before the American Chemical Society, Division of Chemical Literature.⁹

The results of our analyses were somewhat ironic. We found that the mere fact of existence of enhanced resources and communication media does not mean that there will be greater use of available sources of information by those who can benefit from them. Taking the physician-clientele of a medical library network and the industrial clientele of a network scheme such as that put forward in the State Technical Services Act, it is clear that there is a gap between available information resources and their potential users which is best bridged by prosthetic devices such as field representatives who are somewhat akin to the county agent of agriculture. Such representatives have proved to be the most effective means of ensuring the efficient dissemination of useful information, assuming that "efficient" dissemination is that which gets information out to the greatest possible number of potential users, and not merely to the select few who are literature-oriented or who work in institutions which are already richly served with information tools and media. Thus, effective interpersonal communication is likely to remain a basic necessity, and may be needed even more as a result of the enhanced resources resulting from networks.

⁸ Herner, S., et al., A Recommended Design for the United States Medical Library and Information System, 2 v., Washington, D. C., Herner and Company, 1966.

⁹ Herner, S., The Place of the Small Library in the National Network, Journal of Chemical Documentation, 6, 171-3, 1966.

HERNER AND COMPANY

WASHINGTON, D. C. 20037

Director of Information Sciences
Air Force Office of Scientific Research
August 9, 1967

Page 8

Information-Related Equipment

Another small project, also the outgrowth of a previous activity, was concerned with the reduction to written expository form of a series of lectures on non-computer equipment which I gave as part of a course on Instrumentation at the Drexel Institute of Technology Graduate School of Library Science. The final product of this project was a chapter in the proceedings of the Conference on Technical Information Center Administration at Drexel Institute.¹⁰

The essential theme and purpose of this chapter was to define the interface between the older and relatively conventional information-related devices and the more avant garde devices which are beginning to have an impact. A related function of the chapter was to describe and discuss the role of the older devices in the new communication mechanics underlying the current drive toward networks and network concepts. When dissected carefully, it becomes apparent that the new communication mechanics, for the time being at least, is very largely dependent on devices and concepts which have been with us for years. There is, however, a shift, in that old-line devices such as punched tape and microfilm, once contemplated as ends in themselves, are emerging as intermediaries in the processes performed by the newer mechanisms.

One interesting sidelight of the analysis that went into the preparation of the chapter relates to Calvin Mooers' concepts regarding paper tape typewriters and reactive typewriters. Mooers' tape typewriter plan is an example of the great distance that frequently separates theory and practice in this field. On the face of it, the idea of preparing and exchanging punched paper tape records of texts, index records, etc., would seem very attractive and appropriate. However, operationally, there are so many variables to be reconciled (typing styles, codes, available type characters, etc.) that punched paper tape has proved an impossible medium of exchange, and has remained useful only for repetitive typing on the same machine or as input for computers and similar devices.

The concept of exchange has now carried over into the realm of magnetic tape records, still with near-insurmountable problems. For instance, it cost the University of California (Los Angeles) upwards of \$125,000 to do the system design and programming necessary to permit the use of MEDLARS tapes. Unfortunately, the results of the efforts of UCLA are not readily transferable because of the broad range of computer types, configurations within types, and software requirements and conventions.

¹⁰ Herner, S., Composition, Copying, and Semi-Mechanized Filing and Searching Devices, Third Conference on Technical Information Center Administration (TICA 3), pp. 61-73, New York, Spartan Books, 1967.

HERNER AND COMPANY

WASHINGTON, D. C. 20037

Director of Information Sciences
Air Force Office of Scientific Research
August 9, 1967

Page 9

On the other hand, Mooers' concept of the reactive typewriter has proved visionary, and we are now witnessing a rapid expansion of typewriter terminals serving as computer-based repetitive typing devices and as input and output devices for computerized retrieval systems. But this does not involve tape exchanges, and so problems of compatibility do not arise to muddy up the situation.

Reporting Requirements

In further pursuance of the terms of the subject contract, I am enclosing completed DD forms 1473 for each of the publications produced during the applicable period. I am also forwarding, under separate cover, 25 copies of each of these publications.

Very sincerely yours,

HERNER AND COMPANY

Original signed by
Saul Herner

Saul Herner

Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified.)

1. ORIGINATING ACTIVITY (Corporate author) Herner and Company 2431 K Street, N.W. Washington, D.C. 20037		2a. REPORT SECURITY CLASSIFICATION Unclassified	
3. REPORT TITLE STUDY OF DOCUMENTATION PROCEDURES AND MECHANISMS		2b. GROUP	
4. DESCRIPTIVE NOTES (Type of report and inclusive dates) Scientific Final			
5. AUTHOR(S) (First name, middle initial, last name) Saul Herner			
6. REPORT DATE 9 August 1967	7a. TOTAL NO OF PAGES 9	7b. NO OF REFS 10	
8a. CONTRACT OR GRANT NO AF49(638)-1424	8b. ORIGINATOR'S REPORT NUMBER(S)		
9. SUBJECT NO 9769-C1 61445014 4681304	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned) AFOSR 67-2248		
10. DISTRIBUTION STATEMENT 1- Distribution of this document is unlimited.			
11. SUPPLEMENTARY NOTES TECH. OTHER		12. SPONSORING MILITARY ACTIVITY Air Force Office of Scientific Research Directorate of Information Sciences (SRI) 1400 Wilson Boulevard Arlington, Virginia 22209	
13. ABSTRACT This is a review of the work and publications emanating from the subject contract. This work derived from the results of empirical studies performed over a period of years to find solutions to problems of general concern in information science and technology, or of concern to specific clients in government and industry. The work and publications discussed are on the following topics: ASLIB-Cranfield Evaluation Techniques; System Design, Evaluation, and Costings; Review of User Studies; Machine Analysis of Psychiatric Interviews; Functional Aspects of Information Networks; Information-Related Equipment.			

DD FORM 1 NOV 66 1473

Security Classification