SP-2535

SP *a professional paper*

USER REQUIREMENTS, PERSONAL INDEXES, AND COMPUTER SUPPORT

July 25, 1966

Everett M. Wallace

SYSTEM DEVELOPMENT CORPORATION
2500 COLORADO AVE.
SANTA MONICA
CALIFORNIA
90406

SDC

ABSTRACT

This paper discusses the necessity for scientific and technical personnel to maintain personal files and the limitations of centralized information services in meeting the diverse and changing information access requirements of these personnel. A prototype service is described that uses a computer to aid individuals organize, maintain and find what is in their files by means of personalized, printed indexes. Through building a machine-readable record of user indexing practice, the service furnishes a means of identifying current user information requirements and for improved feedback of these requirements to centralized services.

USER REQUIREMENTS, PERSONAL INDEXES, AND COMPUTER SUPPORT

Over the past twenty years an increasing amount of attention has been given to the problems of managing information. Nowhere is the concern more evident or more real than in organizations concerned with scientific research and development. These organizations have led in the development and application of modern data processing technology in the dissemination, indexing, storage, and retrieval of documented information. So far, however, these applications have been confined largely to centralized information services that are, necessarily, limited in their capacity to respond to the immensely variegated, changeable, and time-dependent information requirements of technical personnel. That these limitations are real and troublesome is reflected in the increasing number and results of user studies and in the (so far) rather equivocal attempts to measure and evaluate the effectiveness of these systems and services. Then there is the growing proliferation of small specialized services and centers that, from the viewpoint of managerial control and planning, tend to sprout up like weeds in a well-tended garden. An additional kind of evidence is the prevalence, persistence, and size of personal and office files.

Jahoda, et al., in summarizing previous work, state, "Information gathering habit studies have shown that a significant portion of researchers maintain personal indexes. Studies by Fishenden, Tornudd, and Hogg and Smith, for example, have brought out the fact that 45%...57%...and 66%, respectively of surveyed scientists had and/or used personal indexes. Zwemer has found that nearly every scientist surveyed in a recent study kept a personal file in the way of reprints, abstracts, or notes on cards, and that the average rate of growth of 26 such collections is 330 times per year... In another recent study of the information needs of Department of Defense scientists and engineers, 17% of the interviewed scientists and engineers used personal files as their first source of information, while 51%... relied on their local environment--personal files, departmental files, and colleagues--as a first source of information." (1)

One may conclude from this and other evidence that most of the scientific and technical personnel working in government, industrial, and academic environments need to maintain personal or office files. A smaller but still appreciable number maintain indexes to their files and devote a sizable amount of time and effort to indexing them.

The Problem

The existence of large working files has often been a source of concern to managers because (a) maintenance of individual files seems costly, (b) working files have a tendency toward growth, if unrestricted, and (c) the files reflect individual vocabulary and filing practices and cannot readily be shared by other individuals. Most of the efforts that have so far been directed toward aiding technical personnel in finding and obtaining the information they need has concentrated either upon improvements in centralized services to supplant the need for personal collections and indexes, or upon providing centralized indexing support to reduce the investment of highly paid labor in such efforts.

It seems probable, however, that no matter how centralized information services are improved and expanded in accessibility, scope, and usefulness, there will continue to be a need for building and maintaining personal collections that reflect unique individual requirements and habits of work. Centralized indexing tends to impose a uniform practice and vocabulary that cannot be wholly responsive to individual perspectives and manner of construing subject matter within the context of a specialized set of tasks and goals. There are also entire classes of information and documentation whose interrelations cannot readily be provided for through centralized services.

Consider the following sequence of documentation categories:

working notes
laboratory notebooks
preliminary sketches and drawings
memoranda
formal drawings or diagrams
data
progress reports
test reports

summary technical reports
papers and articles

These are the primary concern of indexing and abstracting services, libraries and information centers

technical orders and specifications
operational manuals
maintenance manuals
trade catalogs

Here we have a sequence in time typical of engineering development projects in which, as we progress from notebooks to memoranda to formal summary reports, detail is lost. Summary technical reports, papers and articles, particularly, tend to be the least informative in detailed technical content. In the course of a project a working engineer will need to correlate information from several of these categories. This leads to a necessarily

idiosyncratic organization of the necessary paper responsive to his changing needs in the course of a task or project. Comparable requirements for the working scientist exist. He too needs information organized in a way that is responsive to the changes in cognitive set that ensue as an inquiry proceeds.

It is improbable that any centralized information service could serve these kinds of transient requirements effectively, if only because they cannot be identified with sufficient precision and in good time. For many years various techniques have been explored to find more effective ways of determining and satisfying the document and information requirements of scientific and technical personnel. Most user studies, however, have depended almost entirely upon the traditional devices of diary, questionnaire and interview plus statistical records of centralized services. This places a heavy reliance on retrospective testimony. Relatively little has been done to study the actual behavior of these personnel. In most organizations it has been difficult to implement effective feedback mechanisms to centralized services on a continuing basis or to arrive at reliable means of measuring the responsiveness of services.

One Approach to a Solution

Over the past three years a project at System Development Corporation (SDC) has been concerned with developing better means of studying user behavior, identifying needs and requirements, and helping individuals improve the organization, maintenance and access to their personal files. Our approach has been to explore the use of a computer in providing printed indexes to individual collections. A service was developed that was intended to meet a diversity of requirements and be adaptable to individual viewpoints, vocabulary and habits of work. An early version of the service has previously been reported. (2)

The current service is called SURF (Support of User Records and Files). SURF is implemented through the corporation's MADAM programming language and system for the IBM 1401. (3) Users of the service index their files, fill out and submit input coding sheets to the service, and regularly receive updated, consolidated indexes. So far in the development of this prototype service, the major effort has been spent in building a tool that would be responsive to real needs and would minimize the effort required by users in coding inputs. This latter consideration is particularly important to the viability of any such scheme. The service must provide for a net saving in time and effort to the individual if he is to use it.

Figures 1 and 2 illustrate, respectively, an input coding sheet and the resultant printed indexes for a bibliographic example. User effort at the input end has been minimized through the use of variable-length fields labeled by field numbers that identify elements within each indexed

entry. Each field number is an Arabic numeral followed by a parenthesis. In Figure 1, field number 1) has been assigned to authors and agencies, field number 3) to subject keys, and field number 2) to elements not to be sorted and alphabetized. The field numbers are automatically deleted from the printed indexes, as illustrated in Figure 2. This kind of free-field input format frees the user from having to concern himself with how his inputs are keypunched and processed. There is no rigid formatting requirement except for identifying the user and the indexed entries.

Ideally, such a service should help individuals find more effective means of organizing and accessing files. Most SURF users have found that they learned to build better indexes through having to use the products of their earlier indexing decisions. Often their initial choices of indexing terms, order of entry, etc., were far from ideal. Through their indexes they could diagnose poor practices and improve their grasp of what is required in nomenclature and perspective. At the same time each user is provided with a product that completely reflects his outlook and manner of representing information. The adaptability of SURF to these ends is indicated by the broad variety of uses to which the service has been put. These uses have included indexes to technical literature and correspondence, engineering data, trade catalog citations, 35mm slides, and a dictionary of grammar rules for programmed query analysis. (4) (5)

A byproduct of the service has been the cumulation of a machine-readable file of user indexing practice reflecting user's needs, perspectives, vocabulary and manner of organizing information vital to their work. Such a file is being used for direct observation of user behavior and needs, and for feedback to centralized services for document description, acquisition, dissemination and retrieval. Analysis of this kind of file has the potential of aiding greatly in a more precise identification and specification of user information requirements. Such analysis offers an additional dimension of study to the information derived from the traditional tools of diary, questionnaire and interview.

Applicability of SURF

The approach of SDC's SURF project has been that of developing successive designs and programs for a small group of customers representing a wide range of different needs, and interacting with them over time to build a truly responsive and individualized service. In a sense this effort could be called an operational experiment. As such it is tied to the particular concerns, personnel and work of one company. It is probable that many organizations would find it advantageous to explore more effective means of serving the idiosyncratic information needs of their scientific and technical personnel in a comparable way. This is not to say that well-managed

individual files will eliminate the need for effective centralized information and document services. Working files and centralized services are complementary types of operations. For any organization there is a desirable balance and a desirable interface between these enterprises. An operational exploration comparable to the SURF development offers one means of identifying ways to improve services while aiding the ability of individual workers to serve themselves.

SURF Input Coding Sheet

Name: Everett Wallace Room 9935 Ext. 6561 Date: 6/24/66

ID Code 1-3	Output Format 4-5	Entry No. 6-9	Card No. 10	Numbered Fields + Field Contents 11-79
WE1	B1	0101	1	2) Jahoda, G. 1) Hutchins, R. D. 1) Galford R. R. 2) characteristics and use of 3) personal indexes maintained by scientists and engineers 2) in one university. 2) American Documentation vol 17 no 2 p71-75 April 1966 1) Florida State U Faculty 2) library
		0079	1	1) Crossley, w. o. 2) The madam system 2) SDC TM-2198/002/00 2) Dec 1965 3) madam programming language and system 3) ibm 1401 programs 2) EW
		0096	1	1) Wallace, E. M. 3) rank order patterns of common words as 3) discriminators of subject content 2) in scientific and technical prose 1) Symposium on Statistical Association Methods for 3) Mechanized Documentation 2) proceedings p225-229 NBS misc pub 269 Dec 1965 3) language processing 2) Doyle

Figure 1. Example of SURF Input

SURF INDEX

FIELD NO. 1 FOR WEI ENTRY NO.

CROSSLEY, W. O. 79
CROSSLEY, W. O. THE MADAM SYSTEM SDC TM-2198/002/00 DEC 1965
MADAM PROGRAMMING LANGUAGE AND SYSTEM IBM 1401 PROGRAMS EW
FLORIDA STATE U FACULTY 101
JAHODA, G. HUTCHINS, R. D. GALFORD R. R. CHARACTERISTICS AND
USE OF PERSONAL INDEXES MAINTAINED BY SCIENTISTS AND ENGINEERS
IN ONE UNIVERSITY. AMERICAN DOCUMENTATION VOL 17 NO 2 P71-75 A
PRIL 1966 FLORIDA STATE U FACULTY LIBRARY
GALFORD R. R. 101
JAHODA, G. HUTCHINS, R. D. GALFORD R. R. CHARACTERISTICS AND
USE OF PERSONAL INDEXES MAINTAINED BY SCIENTISTS AND ENGINEERS
IN ONE UNIVERSITY. AMERICAN DOCUMENTATION VOL 17 NO 2 P71-75 A
PRIL 1966 FLORIDA STATE U FACULTY LIBRARY
HUTCHINS, R. D. 101
JAHODA, G. HUTCHINS, R. D. GALFORD R. R. CHARACTERISTICS AND
USE OF PERSONAL INDEXES MAINTAINED BY SCIENTISTS AND ENGINEERS
IN ONE UNIVERSITY. AMERICAN DOCUMENTATION VOL 17 NO 2 P71-75 A
PRIL 1966 FLORIDA STATE U FACULTY LIBRARY

SURF INDEX

FIELD NO. 3 FOR WEI ENTRY NO.

DISCRIMINATORS OF SUBJECT CONTENT 96
WALLACE, E. M. RANK ORDER PATTERNS OF COMMON WORDS AS DISCRIMI
NATORS OF SUBJECT CONTENT IN SCIENTIFIC AND TECHNICAL PROSE SY
MPOSIUM ON STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUME
NTATION PROCEEDINGS P225-229 NBS MISC PUB 269 DEC 1965 LANGUAG
E PROCESSING DOYLE
IBM 1401 PROGRAMS 79
CROSSLEY, W. O. THE MADAM SYSTEM SDC TM-2198/002/00 DEC 1965
MADAM PROGRAMMING LANGUAGE AND SYSTEM IBM 1401 PROGRAMS EW
LANGUAGE PROCESSING 96
WALLACE, E. M. RANK ORDER PATTERNS OF COMMON WORDS AS DISCRIMI
NATORS OF SUBJECT CONTENT IN SCIENTIFIC AND TECHNICAL PROSE SY
MPOSIUM ON STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUME
NTATION PROCEEDINGS P225-229 NBS MISC PUB 269 DEC 1965 LANGUAG
E PROCESSING DOYLE
MADAM PROGRAMMING LANGUAGE AND SYSTEM 79
CROSSLEY, W. O. THE MADAM SYSTEM SDC TM-2198/002/00 DEC 1965
MADAM PROGRAMMING LANGUAGE AND SYSTEM IBM 1401 PROGRAMS EW
MECHANIZED DOCUMENTATION 96
WALLACE, E. M. RANK ORDER PATTERNS OF COMMON WORDS AS DISCRIMI
NATORS OF SUBJECT CONTENT IN SCIENTIFIC AND TECHNICAL PROSE SY
MPOSIUM ON STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUME
NTATION PROCEEDINGS P225-229 NBS MISC PUB 269 DEC 1965 LANGUAG
E PROCESSING DOYLE
PERSONAL INDEXES MAINTAINED BY SCIENTISTS AND ENG 101
JAHODA, G. HUTCHINS, R. D. GALFORD R. R. CHARACTERISTICS AND
USE OF PERSONAL INDEXES MAINTAINED BY SCIENTISTS AND ENGINEERS
IN ONE UNIVERSITY. AMERICAN DOCUMENTATION VOL 17 NO 2 P71-75 A
PRIL 1966 FLORIDA STATE U FACULTY LIBRARY
RANK ORDER PATTERNS OF COMMON WORDS AS 96
WALLACE, E. M. RANK ORDER PATTERNS OF COMMON WORDS AS DISCRIMI
NATORS OF SUBJECT CONTENT IN SCIENTIFIC AND TECHNICAL PROSE SY
MPOSIUM ON STATISTICAL ASSOCIATION METHODS FOR MECHANIZED DOCUME
NTATION PROCEEDINGS P225-229 NBS MISC PUB 269 DEC 1965 LANGUAG
E PROCESSING DOYLE

Figure 2. Example of SURF Index

References

1. Jahoda, G., Hutchins, Ronald D. and Galford, Robert R. "Characteristics and Use of Personal Indexes Maintained by Scientists and Engineers in one University." American Documentation Vol. 17 no. 2. April 1966 pp. 71-75.

2. Wallace, Everett M. Experience with EDP Support of Individuals' File Maintenance. System Development Corporation. SP-1646. July 7, 1964.

3. Crossley, William O. The MADAM System. System Development Corporation. TM-2198/002/00. December 2, 1965.

4. Wallace, Everett M. SURF: Support of User Records and Files - Description and Operation. System Development Corporation. TM-2912/000/00. April 29, 1966.

5. Wallace, Everett M. A User's Guide to SURF: Support of User Records and Files. System Development Corporation. TM-2913/000/00. June 24, 1966.

Unclassified

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)

System Development Corporation
Santa Monica, California

2a. REPORT SECURITY CLASSIFICATION

Unclassified

2b. GROUP

3. REPORT TITLE

User Requirements, Personal Indexes, and Computer Support

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

5. AUTHOR(S) (Last name, first name, initial)

Wallace, Everett, M.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
25 July 1966	9	5

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
Independent research	
b. PROJECT NO.	SP-2535
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY

13. ABSTRACT

This paper discusses the necessity for scientific and technical personnel to maintain personnal files and the limitations of centralized information services in meeting the diverse and changing information access requirements of these personnel. A prototype service is described that uses a computer to aid individuals organize, maintain and find what is in their files by means of personalized, printed indexes. Through building a machine-readable record of user indexing practice, the service furnishes a means of identifying current user information requirements and for improved feedback of these requirements to centralized services.

DD FORM 1 JAN 64 1473

Unclassified

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Personal Indexes **Computer Support** **File Organization** **Indexes** **Machine-readable records**						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.