SELECTIVE DISSEMINATION OF INFORMATION (SDI)
VOLUME I
PILOT TEST AT U.S. ARMY NATICK LABORATORIES

ARMY TECHNICAL LIBRARY IMPROVEMENT STUDIES (ATLIS) REPORT NO. 15

May 1967

SELECTIVE DISSEMINATION OF INFORMATION (SDI)

VOLUME I

PILOT TEST AT U. S. ARMY NATICK LABORATORIES

Army Technical Library Improvement Studies (ATLIS)
Report No. 15

14 May 1967

Prepared for

U. S. Army Natick Laboratories
Natick, Massachusetts 01762

Contract No. DA 19-129-AMC-957 (N)

by

Information Dynamics Corporation
80 Main Street
Reading, Massachusetts 01867

IDC Report No. 8074-2

PAGES _____

ARE

MISSING

IN

ORIGINAL

DOCUMENT

ABSTRACT

This report presents the results of a nine-month test of a prototype SDI system developed for Army Technical Libraries. During the pilot test, one thousand documents were cataloged and indexed and disseminated to twenty-five scientific and technical personnel at U. S. Army Natick Laboratories. During the course of the pilot test, statistics were accumulated on operating costs and various parameters affecting the relevancy of disseminated information. The prototype SDI system utilizes graphic techniques for producing multiple copies of each document abstract. The output of the computerized matching function is a set of EAM cards which are subsequently used to collate reproduced abstracts for dissemination. Collation is accomplished by interspersed gangpunching, merging, and sorting techniques.

FOREWORD

Over the past decade, libraries have gradually assumed a more active role in making recorded information more available to the technical community. Through the assimilation of data processing techniques previously relegated to scientific computation and accounting, the library has been able to decentralize access to its collection of stored information. Selective dissemination of information is an extension of this trend towards decentralization of information access. An SDI system delivers, to the users in the field, a representative portion of a document selected on the basis of an interest profile stated by the user.

During a first phase of this contract, a number of representative SDI systems was studied and described in a separate report. This report is available from Defense Documentation Center as AD 636916. It compares the various implementation methods in current use.

Using the results of the study conducted in Phase I, the prototype SDI system was developed for the Army Research Community. This report gives the operating results obtained from a nine month pilot test of the prototype system. The pilot test demonstrated the utility of the system in starting up an SDI program at U. S. Army Natick Laboratories. Other laboratories may utilize the system as it exists directly or may choose to modify it to conform with on-going automated library services.

This final report was prepared for the ATLIS Program under the direction of Mr. Robert Martin, Chief Librarian, Army Natick Laboratories, Natick, Massachusetts. All work was performed by Information Dynamics Corporation, 80 Main Street, Reading, Massachusetts under contract DA19-129-AMC-957(N).

ACKNOWLEDGMENTS

The very nature of a pilot test implies that new ground is being broken. The solution to problems invariably encountered in a pilot test requires the concerted efforts of many heads and many hands. To Mr. Robert Martin and his staff, Mr. Robert McDonald and Mrs. Edna Birmingham, I am indebted for receiving the useful suggestions and cooperation necessary for the success of the project. I am also grateful for the cooperation of Mr. Ronald Geromini, Director of the NLABS Computational Facility, and especially for the aid given me by Mr. Frank Moran in overcoming the hurdles of programming for the GE 225.

The success of this project can be, to a great measure, attributable to the competency of Mrs. Antoinette DiPerna, who indexed most of the documents processed in the pilot test. But in the final analysis, thanks are owed to Mr. James Flanagan for his able assistance in managing the project and to Miss Charlotte Dziurkiewicz and the IDC secretarial staff who "made the thing work."

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

TABLE OF CONTENTS (cont'd)

TABLE OF CONTENTS (cont'd)

LIST OF TABLES AND FIGURES

LIST OF TABLES AND FIGURES (cont'd)

I. INTRODUCTION - PURPOSE AND SCOPE OF THE PILOT TEST

The overall objectives of this contract were: to survey a representative sample of existing computerized or semi-mechanized systems for selective dissemination of information (SDI); to develop and describe an effective procedure for establishing and operating a model SDI system; and to perform a pilot test of the recommended system and evaluate its suitability for implementation at Army Natick Labs (NLABS) and other Army research activities. The work was accomplished in two distinct phases, separately funded and sequentially conducted.

Phase I of the project encompassed the survey of existing systems and the design and description of a model SDI system incorporating the best features of existing systems and any innovations that would enhance its efficiency or economy. Operating SDI systems were studied which demonstrated different techniques for accomplishing selective dissemination of information. Operational aspects examined were: the type of publications constituting information input; the basis for screening and selecting input and levels of personnel performing screening or selection; the method of abstracting and indexing; the method of vocabulary control; the types of personnel comprising the user group; the number of users and their subject interests; the method of deriving a users' interest profile; the strategy for matching interest profiles against document descriptions; the nature of the notification sent to the user; the type of feedback collected; statistical controls exerted; and statistics reported on the relevancy of notices, distribution of notices, and operating costs. At the conclusion of Phase I, the results of the systems survey and the description of a model system were included in the final report, "Selective Dissemination of Information: Review of Selected Systems and a Design for Army Technical Libraries" (AD 636916) August, 1966.

The objective of Phase II, results of which are here reported, was to pilot test the system designed and recommended in Phase I. The pilot test was conducted with twenty-five professional personnel at Army Natick Laboratories (NLABS) and included the processing of 1000 technical information units selected from serial and monographic publications or abstracts.

Several sub-objectives of Phase II were implicitly stated in the Phase I final report and were, by contract, adopted as objectives of Phase II. One of these objectives was to test the feasibility of implementing an SDI system which started at time "zero" with no controlled vocabulary and built a controlled vocabulary as a by-product of operation. Another very important objective was to evaluate as accurately as possible all costs associated with the operation of the pilot system. These costs were to include labor, materials, special equipment, machine utilization, and space requirements but were to exclude the acquisition costs of information and the cost

of supplying full-size documents or microreproductions requested as the result of SDI operations.

The methods utilized in the pilot test were intended to be capable of extrapolation to other Army Research Activities, and the system was to be described in sufficient detail for direct implementation without outside consultation. Procedures for implementation are described in Volume II of this final report.

II. THE USER POPULATION AT NLABS

A. INTEREST GROUPS REPRESENTED AT NLABS

The most characteristic attributes of Army Natick Laboratories are its extreme diversity of activities and the great depth of detail represented by each activity. Probably there is no comparable assemblage of this variety of talents at any other military installation. An equivalent range of interests could only be expected to exist at some of the largest universities, such as M. I. T. or the California Institute of Technology. The findings of this study indicate that the variety of activities at NLABS imposes more stringent requirements on the design and pilot test of an SDI system than would be encountered at any other Army installation.

1. The NLABS Mission

In brief, the NLABS mission is to support personnel in a hostile environment. Support functions include provisioning and protection against the elements, the implements of warfare, and attack by animal and vegetable predators. They also include general supply -- food, fuel, clothing, personal equipage, shelter, and weapons. Environmental conditions range from those encountered below the surface of the sea to outer space and in latitude from the poles to the equator. Activities range from basic research in improving the properties of materials to the styling and tailoring of military uniforms.

2. Organizational Structure

Army Natick Laboratories presently conducts research in the following seven divisions:

Air Drop Engineering

Clothing and Organic Materials

Containers

Earth Sciences

Foods

Mechanical Engineering

Pioneering Research

These are supported by a technical services division and administered by a Commanding General, Deputy Commander, Scientific Director, and supporting functions associated with base operation.

The Air Drop Engineering Division develops systems for in-flight delivery of personnel, equipment, and supplies. The division studies materials suitable for container fabrication, parachutes, harness straps, and packaging, and it designs and tests parachutes and other air delivery devices.

The Earth Sciences Division measures and describes world-wide climatic conditions that would affect the requirements for human support. Geographers, meteorologists, and cartographers develop statistics, maps, and guides covering all types of environment from jungle to mountain and desert to arctic regions.

The Food Division conducts basic and applied research in the development of stable rations and methods for producing food and potable water where these are lacking in the environment. Food technologists are concerned with the affects of temperature, humidity, radiation, and biological and chemical contamination on the storage properties, flavor, appearance, and consistency of foods and beverages.

The Clothing and Organic Materials Division designs a full range of clothing for both work and dress, and it studies factors affecting comfort, serviceability, wear, launderability, appearance, heat dissipation, and freedom of movement. In addition, the division studies systems of materials which can be combined for protection against projectiles, anti-personnel mines, and human traps.

The Pioneering Research Division conducts basic research in materials development, and it studies the effects of radiation, biological and chemical attack, and other environmental factors on various natural and synthetic materials.

The Container Division develops new packaging materials and methods for forming and handling packages. Packaging must be developed which meets current military specifications and special additional requirements imposed by the environment and the circumstances under which the package is to be used. In addition, the Container Division develops specifications for new types of packaging.

The Mechanical Engineering Division investigates the properties of materials and supports other divisions in developing materials for specialized purposes.

Throughout all divisions there is a strong interface with electronics because of the dependency on modern instrumentation and because of the interface between man and machine which affects the design of personnel equipment.

3. Pure Research

Pure research is conducted primarily in the development and characterization of new materials and products for human consumption and protection. Research is under way in the development of new methods for producing foods and new materials for protecting the physical integrity of the human body and for maintaining its chemical and thermal equilibrium. Basic research is constantly conducted on all environmental factors which affect both the human and materials which he uses. In the social sciences, psychologists are at work to determine the factors which control the efficiency and comfort of man and the various psychological factors that influence the utility and effectiveness of personnel equipment.

These representative research projects at NLABS indicate the variety of problems approached. A complete listing would serve no additional purpose.

4. Applied Research

Professional personnel at NLABS are primarily charged with the responsibility to develop reliable equipment and supplies which meet the requirements of a particular mission or battlefield condition. Almost all development activities ultimately lead to a usable end item such as a uniform, tent, POL storage facility, air drop delivery system, or field ration. Much of the applied research is therefore pointed towards the testing and evaluation of materials as they appear on the market or as they are developed in the research laboratory. Extensive testing is performed on various plastics, synthetic fibers, fabrics, leather, protective face creams, fiber-reinforced materials, elastomers, and other products going into the fabrication of an end item.

Complete clothing and personnel equipment systems are also designed, fabricated, and tested at NLABS facilities. Complete clothing outfits, bedding, tentage, and field packs are tested extensively under simulated environmental conditions to evaluate them in terms of serviceability and comfort.

5. Research and Development Facilities

The main research buildings at NLABS consist of climatic chambers, a complete shop, a radiation laboratory, subsistence evaluation laboratory, and individual laboratory offices. A gigantic solar furnace is used to test the high-temperature properties of materials. Each office laboratory is equipped to accomplish the objectives of individual projects. There is an abundance of all types of chemical analytical equipment and materials testing equipment. Equipment is also available for forming and fabricating various materials on a laboratory and semi-production basis.

6. End Products and Services

Although NLABS does not manufacture an end item on a production basis, it does provide extensive models and feasibility studies in anticipation of full-scale production. Primarily, NLABS is an experimental design facility which turns out one-of-a-kind items that are subsequently mass produced by contractors. In many instances, NLABS collaborates with other military and non-military Government research activities which require utilization of the specialized knowledge, skills, and equipment available at the NLABS campus.

7. Summary of Interests

The above project facilities and organizational descriptions are given to acquaint the reader with the extreme variety of interests to be served by an SDI system at Army Natick Laboratories. Some of the research fields encompassed by NLABS personnel are mathematics, physics, chemistry, meteorology, anthropology, chemical, electrical and mechanical engineering, biochemistry and bacteriology, psychology, nutrition, geography, micology, geophysics, entymology, parasitology, radiology, process control quality assurance, and metallurgy. Of approximately 600 professional personnel, thirty percent hold a Master's or Doctoral degree.

B. SELECTION OF PARTICIPANTS IN THE PILOT TEST

That all professional personnel at NLABS could not be included in a pilot test was immediately obvious. This observation necessitated selecting a sample of participants that would not impair the meaningfulness of the pilot program but would represent the diversified user population at NLABS.

1. Objectives of Selection

The primary objective of selection was to limit the number of participants to a manageable group which preserved a depth and breadth of subject interests characteristic of the entire NLABS population. Individuals who represented a wide number of subject interests were thus selected. Some had an in-depth interest in a particular subject, and others had a shallower interest in a wide number of topics. An additional requirement was that lines of communication be established between the project and the participants so that adequate feedback could be obtained on the effectiveness of dissemination achieved by each SDI cycle.

It was ultimately determined that the above requirements could be met by a sample population containing twenty-five participants. As a result of an analysis of the activity of the professional personnel, the Clothing and Organic Materials Division was selected as a pilot test bed. Within this division activities ranged from pure research to finished product development, and a broad range of subject interests were represented.

2. Range of Subject Interests

The range of subject interests represented by the twenty-five participants in the pilot test is shown in Table I.

3. Information Acquisition Habits of Participants

The selection of participants in the pilot test was balanced with respect to current information acquisition habits. Some participants were extremely heavy users of the technical library, whereas others used the library very infrequently. Some of the participants read extensively after working hours, whereas others utilized recorded literature very little. Some had comprehensive literature collections at their own desks, whereas others had only a few publications on hand.

TABLE I
SUBJECT INTERESTS OF PILOT PARTICIPANTS

Acetylenic organometallic compounds	Life support systems for terrestrial, space, and underwater applications
Blast and shock wave phenomena	Metal cleaning and descaling compounds
Clothing manufacturing processes	Metal finishing
Color measurement	Military uniforms
Criteria for human comfort	Molecular structure and molecular spectroscopy of polymers
Cryogenic life support systems	Noise and impact energy attenuation
Decomposition of sulfonium salts	Permeability of fabric systems
Deterioration of synthetic fibers	Physical and chemical properties of polymers
Effect of fabric geometry on physical properties	Polymer rheology and molecular structure
Foam-in-place plastics and low-temperature foaming systems	Protective clothing
Fluid mechanics	Protective finishes for textiles
Fuels	Reproduction supplies
Gel permeation chromatography	Soaps and detergents
Graft copolymers	Synthesis of monomers for polycarbonates
Heat and mass transport through films and fabrics	Synthetic fibers
High-impact performance of fibers and fiber assemblies	Techniques for testing and examining fabrics
High-temperature properties of materials	Textile dyeing and finishing
High vacuum equipment and techniques	Textiles for clothing
Impact behavior and abrasion resistance of textiles	Thermal and nuclear radiation protection
Ionic polymerization	Water repellency of fabrics

Information attitudes of the users were also somewhat balanced. A few users were extremely discriminating in the type of information that they would read, but a few were satisfied by a wide variety of topics at varying depths of detail. On the average, however, the reading habits of the sample population followed the norms which one would anticipate in the entire NLABS population. Both extremes and median points in each analytical category were represented in the sample group.

A characteristic of the group that was thought to be representative of NLABS personnel was the degree of reliance on the various categories of literature. The user group relied most heavily on journal literature and least on Government research reports obtained through secondary Government publication sources such as CFSTI, NASA/STIF, and DDC. At least half of the group, however, received Government research reports on direct distribution from other Governmental laboratories and contractors.

III. INFORMATION RESOURCES AT NLABS TECHNICAL LIBRARY

A. COLLECTION COMPOSITION

Current conservative estimates for NLABS holdings are 1000 periodical titles, 15,000 monographs, and 75,000 technical reports. Figures for the journal and technical reports collections do not include a separate file of microfilm procured from various Government, university, and commercial sources. In addition to the literature collection, NLABS has an excellent collection of maps to support its terrestrial sciences research.

1. Journals

The mixture of activities in both basic and applied research is most exemplified by the NLABS journal collection. At least one-fourth of the journals received regularly are trade publications which describe the latest application of new materials, processes, manufacturing techniques, and product developments. The major portion of the journal collection contains descriptions of the latest scientific advances in materials research, food technology, synthetic fiber production, plastics, rubber, and other technological subjects. Each year NLABS publishes a list of the technical papers delivered or published by its scientists. Many of these papers appear in the very journals to which the Library subscribes. As a matter of interest, the abstract of one of the these papers was selected by the SDI system for delivery to its own author, who was one of the twenty-five participants in the NLABS SDI pilot program.

The major fields of concentration of the NLABS journal collection are organic and inorganic chemistry, biological and medical sciences, behavioral and social sciences, and radiology, and to a lesser degree, the engineering sciences.

2. Books

The 15,000 volumes of books are stored on open shelves accessible to the library users as are the current journal subscriptions and 25,000 bound journal volumes. The book collection is concentrated in subject areas similar to those represented in the journal collection. The books are completely cataloged by the Library of Congress system, and each book is represented by cross-filed card sets in the library card catalog which is open to the users.

3. Technical Reports

The 75,000 technical reports include approximately 5,000 laboratory notebooks and classified reports. At the present time, laboratory notebooks are being cataloged by subject. Technical reports authored

by NLABS scientists are completely indexed and multiple access may be gained to them through the library card catalog. Reports received automatically from other Government laboratories are not completely indexed, however. Most of the technical reports which the library orders as a result of a user request are allowed to remain with the user.

B. COLLECTION UTILIZATION

Over 60% of the personnel who would have occasion to utilize the library are considered regular users. The majority of these regular users are scientists and engineers, not administrators. At any time during the working day, one can find someone busily at work in the library. Since unbound journals are not circulated, a high proportion of on-site utilization can be attributed to usage of the journal collection. Fifty percent of the book collection is circulated annually. Requirements for return of a borrowed item are not stringent unless it is needed by another user.

The technical reports collection, 75% of which is authored by NLABS scientists, receives moderate use. Only 2% of the reports are circulated annually. The low use rate of technical reports can be attributed to an efficient initial distribution system and also to the availability of personal copies from such agencies as Defense Documentation Center and National Aeronautics and Space Administration. Microfilm plays a negligible information transfer role as do classified technical reports in hard copy. Special journals and articles are obtained in microform when high cost and limited use do not warrant acquisition of full-size copy.

Users are permitted to remove current journal issues temporarily to copy selected articles on off-site office copy equipment. These copies form a nucleus of loose materials stored by many scientists within their own offices. They are often supplemented by copies obtained through various university copy services. Most of the scientists have their own collections of books, some of which are quite extensive.

C. INFORMATION PROJECTS

The library periodically publishes an accessions list which announces books and technical reports. Sixteen-hundred reference requests are answered annually, some of which occasion the use of the Defense Documentation Center bibliographic service. Occasionally journal articles or technical reports are translated by members of the staff or translations are sought from other NLABS scientists. A title/author/subject card catalog is maintained for books, periodicals, and technical reports authored by NLABS scientists.

Efficient use is currently being made of most of the NLABS collection. The area which could have the greatest increase in utilization is the technical reports collection. Minimal use is made of the abstract bulletins published by such Government agencies as DDC, NASA, and AEC. Although low utilization of the technical reports collection can be attributed to primary distribution, much information still remains hidden within these announcement media which could be disseminated to advantage.

Although the journal collection is heavily read, many of the participants complained that they had insufficient time to read all of the journal literature that they required. Thus, an SDI system could enhance the present high utilization of the journal literature by supplying the reader with sufficient information for him to select the most pertinent articles.

IV. SDI PILOT SYSTEM DESCRIPTION

A. INTRODUCTION

One of the objectives of this program, as stated in the introduction, was to develop a model system that contained the best elements of successful systems. This objective was accomplished in the model by combining techniques utilized in the fields of digital and graphic information processing, in addition to normal SDI techniques.

A novel aspect of the proposed and tested search strategy in this model is that matching user interest against document description is accomplished by a combination of weighted-term and Boolean strategies. Each method can be used separately, or in combination, without program modification or change in operating procedures. This facility presents a unique opportunity to adjust match strategy on the basis of individual experience.

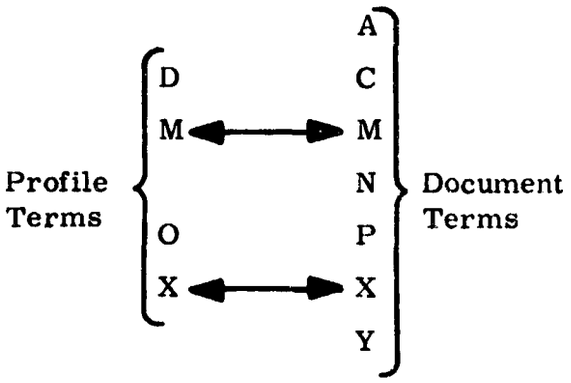
An additional feature, graphic reproduction of the SDI notice, is novel in method, but not in scope. Other systems have produced multiple copies of abstracts on an office duplicator or printing press, but none have mated the selective collation capability of EDP equipment with the speed and accuracy of duplicating equipment. In this system, information is selectively collated by EDP equipment, but reproduced by conventional reproduction equipment. This feature preserves the capability to reproduce whatever the eye sees, and yet retains selective collation of printed abstracts by punched-card sorting techniques.

B. THE INFORMATION/INTEREST MATCH STRATEGY

The Match Strategy is that criteria used for associating an information unit with a user profile. The variables in this two-valued function are an abstraction of the user's interests and an abstraction of the contents of a unit of information (a book, journal article, technical report, etc.).

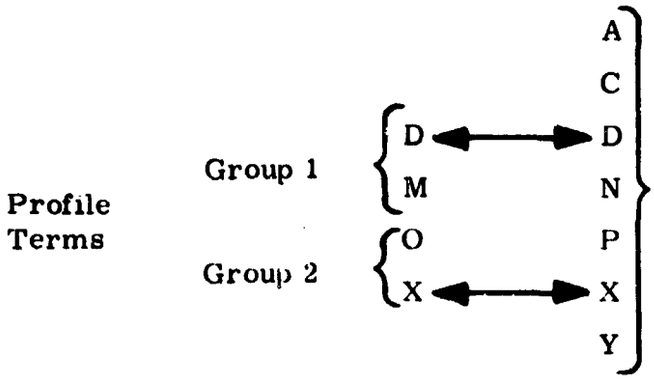
Three basic match strategies were isolated during Phase I of this contract: percentage-of-matching-terms (See A in Figure 1), Boolean (See B in Figure 1), and weighted (See C in Figure 1). The first of these made no use of what can be called the multidimensional aspect of information retrieval (i.e., each dimension is represented by a set of terms describing a different facet of a question: retrieval can be sharpened by recombining the different facets, in a controlled manner, into a single search query). Since the second and third strategies did employ a multidimensional approach, they showed greater promise of yielding more relevant selections. The comments that follow reflect the latter two methods plus a third, a combination of the Boolean and the weighted.

A) PERCENTAGE-OF-MATCHING-TERMS STRATEGY



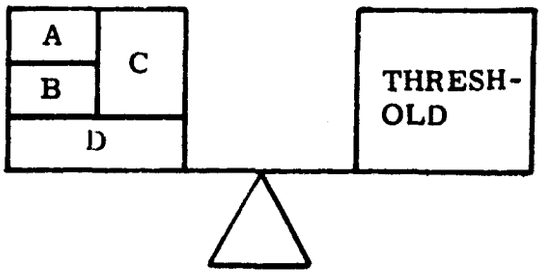
50% of profile terms match document terms.

B) BOOLEAN STRATEGY



At least one term in each profile group matches a document term. Boolean statement is (D or M) and (O or X).

C) WEIGHTED STRATEGY



Profile terms A, B, C, and D match document terms AND sum of weights equals or exceeds threshold.

Figure 1 Information/Interest Match Strategies

1. The Boolean Match Strategy

In the Boolean Match Strategy, the user's interest is represented by one or more sets of descriptive terms, but the information content of a document is represented by only one set of terms. A selection results if the set of document terms intersects with every set of terms describing the user's interest (See B in Figure 1). The user's profile thus takes on the appearance of a truth statement such as, "if (a or b or c) and (d or e), then select," where a, b, and c constitute one term set and d and e constitute a second term set, each describing a different facet of the same user interest.

To implement this search strategy in this pilot test, each set of terms describing a facet of a user's interest was assigned a unique number (See Figure 2). The numbers were used merely as labels for the different sets and did not indicate degree of importance or order of preference. The terms in each profile are ordered by set number and compared serially by computer against the terms describing each document. Each time a user term matched a document term, the event was marked by setting a computer switch. When a set-number change was detected, the state of the switch was tested. If the switch was off, the document was rejected. If the switch had been on, then comparison continued with the next group of terms in the interest profile. If the switch was on for the last set of terms, then the document was selected.

2. Weighted Term Match Strategy

In this strategy, each term in the user profile is assigned a number whose value is chosen by the user. The numbers (called weights) may be chosen by ranking the terms in order of importance. The weights associated with profile terms which match document terms are summed together. If the sum of the weights exceeds a threshold, also stated by the user, the document is selected (See C in Figure 1).

In the original concept, the value of the weight was to be proportional to the strength of the user's interest. In several systems currently operating, weights can range from 1 through 9. Since the total number of terms in any given profile and the pattern of weights are unique for each person, every profile may have a different threshold. All profiles, however, can be normalized to a threshold of one by dividing the user's stated threshold into each of the weights assigned by him (See Figure 2). For example, if a user assigns a weight of 9 to one term and a weight of 1 to all other

IDC MATCH FUNCTION

Profile Term	Weight	Normalized Weight
SET NO. 1		
Geometry	2	2/15 = .133
Weaves	4	4/15 = .266
Fabric Construction	7	7/15 = .467
SET NO. 2		
Bullet Proof Fabrics	1	1/15 = .067
Denims	5	5/15 = .333
Broadcloth	7	7/15 = .467
Duck	4	4/15 = .266
Jersey	7	7/15 = .467
Knit Fabrics	5	5/15 = .333
Mesh Fabrics	5	5/15 = .333
SET NO. 3		
Abrasion Resistance	4	4/15 = .266
Flexibility	6	6/15 = .400
Slip Resistance	2	2/15 = .133
Stiffness	6	6/15 = .400

THRESHOLD = 15

NORMALIZED THRESHOLD = 15/15 = 1.000

Figure 2 Pilot Test Profile

terms in his profile, and requests a threshold of 10, then a normalized profile would show a weight of .9 for one term and .1 for the remaining terms. By normalizing to a unit threshold, the threshold value of 1 can be automatically supplied by the computer program rather than by the profile.

In the pilot test, the normalization to a unit threshold was adopted. Weights as stored in machine-readable data ranged from 0 to .999. These weights were summed algebraically and compared against a threshold of .999. If the sum of the weights equaled or exceeded this threshold, the document was selected.

3. Combined Boolean and Weighted Term Match Strategies

In the pilot test, the Boolean and Weighted Match Strategies could be combined, at will, to increase the discriminating capacity of selection. Actually, both weighted and Boolean aspects were present in all profiles. But the assignment of weights and the choice of sets of terms determined whether one or the other or both methods were in effect. For instance, a profile containing only one set of terms acted as if it were a straight weighted profile, since a selection was limited only by the sum-of-the-weights criterion (See A in Figure 3). Conversely, a profile consisting of several sets of terms, each having a weight of .999 would act as a straight Boolean strategy, since selection was limited only by the Boolean criterion (See B in Figure 3). A combination of the two strategies was operable when the profile contained several term sets having weights less than .999 (See C in Figure 3).

4. The Logical Negative

In combining the Boolean and weighted strategies, the logical negative was implemented by assigning a weight of $\neq 0$ to the negative term. If any profile term having a weight of $\neq 0$ matched a document term, the document was de-selected. That is, the document was barred from selection, even though it may have met both Boolean and weighted criteria for selection.

5. Non-Subject Control Over Dissemination

During the course of the pilot test, it was found desirable to include some controls over dissemination which were not subject oriented. One of these which had been foreseen, but not implemented because of the limited document input, was security classification. In a full-scale

A) WEIGHTED STRATEGY

Profile

Term	Weight
A	.100
B	.300
C	.100
D	.800
E	.800

THRESHOLD = .999

Match will occur for
A + B + D
A + C + E etc.

Match will not occur for
A + B
A + B + C
B + C etc.

B) BOOLEAN STRATEGY

Profile

	Term	Weight
SET NO. 1	A	.999
	B	.999
	C	.999
SET NO. 2	D	.999
	E	.999

Match will occur for
A + D
A + E
A + B + D etc.

Match will not occur for
A + B
A + C
D + E etc.

C) COMBINED BOOLEAN AND WEIGHTED STRATEGIES

	Term	Weight
SET NO. 1	A	.100
	B	.300
	C	.100
SET NO. 2	D	.800
	E	.800

Match will occur for
A + C + D
A + C + E etc.

Match will not occur for
A + D, C + D etc.

Figure 3 Pilot Test Match Strategy

system involving the processing of classified information, it would be necessary to limit dissemination according to security classification of document and security clearance level of user.

An unanticipated requirement for non-subject control over dissemination was the participant's desire to be screened from levels of subject coverage that were below or above his present requirements. In the future, some mechanism should be provided for describing a profile interest in terms such as superficial, intermediate, or detailed. Subject coverage alone is not a sufficient criterion for selection. A result of not having such a provision in indexing both profiles and documents was that documents matched a user's interests by subject, but not by depth of coverage. Thus, someone who had only a superficial interest in a topic might be supplied with information of a highly-detailed nature or vice versa.

Other non-subject categories might play an equally important role in selection or, conversely, in screening. For instance, such categories as materials application, material identification, and properties of materials might be devised to screen out unwanted information, even though it might be selected on the basis of a term-for-term match. This criteria was not necessary in the pilot test because of the limited number of documents processed and the relatively small number of notices resulting from selection. In a full-scale system, however, the selection process will need to be more discriminating to avoid swamping the reader with a large number of peripherally interesting selections.

Early in the pilot test, an experiment was performed to determine the feasibility of indexing documents by other-than-subject terms. In the experiment several documents were indexed by contract numbers, authors, publishers, corporate sources, and other non-subject indicators. Each type of indicator was given a distinctive glossary number similar to those assigned to ordinary subject terms. A demonstration proved that indexing by non-subject terms poses no additional problems to the mechanics of matching and selection. Non-subject terms in profiles could be dispersed in any set of terms comprising the subject portion of the interest profile. The fact that glossary numbers differ does not affect the selection process, since there is no program requirement that glossary numbers assigned to terms in a profile be identical either within the set of terms or within the entire profile.

Non-subject terms may have more value when used in the negative sense than when used in the positive sense. Several participants spoke a foreign language fluently and many had knowledge of one or more

foreign languages. However, some participants neither spoke nor read a foreign language; foreign-language abstracts would be of no direct use to them. Thus foreign language designation could be used in these cases to screen out selections that would otherwise be made on the basis of subject alone.

Several participants expressed a desire to exclude all articles from specific trade journals, either because of the usual poor quality of their articles or because of the lack of thorough coverage in them. Such exclusions would be feasible, as would those based on contract numbers, authors, corporate sources, and publishers.

Exclusions could also be based on other factors: the participant may cover specific sources through his own regular reading; or he may already be informed of certain activities through prior experience.

C. INFORMATION SELECTION AND PREPARATION FOR DISSEMINATION

Due to the wide variety of interests exhibited by the participants in the pilot program, and to the large number of subjects represented by the document collection, it was necessary to screen the document input to assure a reasonable number of selections in the pilot program. Every document selected for inclusion in the test sample underwent the processing steps outlined below. Documents matching user interest profiles underwent additional processing steps.

1. Information Selection - Preliminary Screening

The purpose of preliminary document screening was to make possible the selection and dissemination of a sample of documents in the time limit imposed by the contract. During the pilot program, 1000 documents were to be indexed and matched against the interest profiles for twenty-five Natick Laboratories participants. If the sample of 1000 documents were selected at random from the total population of information resources available at the NLABS library, a relatively small percentage of these documents would result in selections because of the extreme breadth of interests represented within these resources.

The total NLABS current information resources were subject to both screening and selection processes. In the screening process, journals containing a low percentage of relevant information were eliminated from further consideration. These journals were oriented primarily

toward the food technologists, leather chemists, and other fields not represented by the interests of the participants. After screening, the current journals selected for further processing were examined for specific articles to be included in the test sample. In addition, articles were selected as represented in the Technical Abstract Bulletin published by Defense Documentation Center and Scientific and Technical Aerospace Reports published by the NASA Scientific and Technical Information Center. Originally books were scheduled for inclusion in the test sample, but they were dropped from consideration because of the relatively small number of acquisitions received in the subject area during the pilot test.

Articles were selected on the basis of their potential applicability to the interests of the participants. There was no direct connection made between an article and a participant so that there would be no guarantee that a match would occur even once. Rather, an article was selected because it related, in some way, to the interests of the participant group. The selection process merely increased the probability that a document would be selected at least once. If the documents had been selected at random, perhaps only 10% or less would have matched any of the participant's profiles so that 90%, or 900 documents, would have been indexed and processed without any selections taking place.

There is some concern that the selection process might have biased the sample, so that a higher percentage of selections would be relevant than would have been from a random sampling. To a very small degree this hypothesis is true because subject terms applied to documents were limited to the subject content of the samples selected for inclusion in the pilot test. However, the relationship between match strategy discrimination power and the breadth of subject coverage is not at present known, whereas the undesirable effect of processing a large number of documents yielding a relatively low number of selections was quite predictable.

2. Information Extraction

After selection, each article underwent subject indexing and descriptive cataloging. In descriptive cataloging, journal citation and article title were recorded in machine-readable form.

Three different subject heading lists were used in subject indexing. These were: Thesaurus of Engineering Terms, Thesaurus of ASTIA Descriptors, and Textile Technology Terms - An Information Retrieval Thesaurus. The ASTIA Thesaurus was supplemented by DDC Descriptor Frequencies and the ASTIA Chemical Thesaurus. The various thesauri were used to index both documents and interest profiles in the subject areas to which they were best suited.

During the project, four indexers were assigned to the document indexing tasks. This procedure precluded the possibility that the indexer was subconsciously steering an abstract to a participant, since three of the indexers were totally unfamiliar with the interest profiles. Assignment ranged from three to thirty terms per document with an average of eight. All terms were put in machine-readable form.

3. Information Preparation for Dissemination

The standard processing for each article consisted of the following steps: (1) select journal article, and mark journal table of contents, (2) Xerox journal table of contents and title page of each article marked, (3) assign accession number and stamp on table of contents (See Figure 4), title page, descriptive cataloging keypunch form (See Figure 5), and subject indexing keypunch form (See Figure 6), (4) perform subject indexing and descriptive cataloging, (5) cut article title page for reproduction (See Figure 7), and (6) keypunch title, journal citation (See Figure 8) and term cards (See Figure 9).

In preparation for the first SDI cycle, title pages were not cut for reproduction. As a result of operating the first cycle it was discovered that approximately 50% of the documents processed were selected at least once. As a result of this discovery, a decision was made to cut all title pages prior to exercising the match function, so that articles would be available for immediate reproduction.

D. VOCABULARY CONTROL

As a result of an analysis performed in Phase I, a decision was made to utilize a controlled vocabulary as opposed to a vocabulary of terms selected from free text. This decision was based on the fact that results were generally better for those systems using a controlled vocabulary.

1. The Multiple Glossary Concept

Since a technical thesaurus, in the areas of textiles, plastics, high-temperature properties of materials, etc., was not available in a single source document at the commencement of the pilot test, it was necessary to utilize several of the existing glossaries, thesauri, subject headings lists, and other authorities currently available from individual sources. On the basis of a preliminary analysis of the interests of the participants, several existing authority lists were found to be applicable. These were:

QUARTERLY REVIEWS

1966, Vol. XX, No. 3

LONDON:
THE CHEMICAL SOCIETY

Figure 4 Xerox Copy of Journal Table of Contents

Influence of Oblique Channel Ends on Screw-Pump Performance

Flow of very viscous liquids through screw pumps and metering zones of extruders has been investigated by various authors, a few of whom are listed in the references (1-6). Screw-pump equations, which relate flow rate to speed, viscosity, and discharge pressure, are derived in these papers for a number of simplifying assumptions.

The helical screw channel is replaced by a straight prismatic channel with a rectangular cross section in all but one theory (5, 6). In the simplified theory, one assumes such a large aspect ratio, i.e., ratio of channel width to depth, that flow rates can be [illegible] from well-known velocity distrib[illegible]

Congruent [illegible]

...plified theory (2, 3, 4) that are used in the derivation of end effects are discussed first. A schematic diagram of a screw pump is shown in *Figure 1*. The more important nomenclature is shown in the cross section of *Figure 2*. The screw-pump channel is formed between the flights, the screw root surface, and the barrel. Note that the channel has oblique ends, which are ignored in the simplified theory. Isothermal flow of a very viscous Newtonian liquid is assumed.

The flow analysis is simplified when velocities are described relative to the screw. These relative velocities do not change when either the barrel is at a standstill and the screw rotates, or the screw is at standstill and the barrel rotates with an equal speed in the opposite direction, provided inertial and body forces can be ignored. Thus the theory is restricted to low Reynolds numbers, which requirement is satisfied in most screw pumps since they are exclu[illegible] pump very viscous liquids at l[illegible] the

$$\frac{\partial}{\partial y} = \mu \left[\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 v}{\partial z^2} \right]$$

$$\frac{\partial p}{\partial z} = 0 \qquad (3)$$

Congruent velocity dis[illegible]ibutions in parallel cross sec-

	1	3	9	11	17	20	24	80
[illegible]	1 2	00 00[illegible]	0 1	0 1 0 1 0 7 0 1 2 2 [illegible] 3 2				
TYPE 2	1 2		0 2	title				
TYPE 3	1 2		2 1	source				

Polymer Engrg & Sci 7/1 Jan 67

Figure 5 Xerox Copy of Journal Article Title Page With Keypunch Form and Journal Source

DOCUMENT DATA WORKSHEET
PART B: Terms

1 8

J.2 0,0,0,7,1,2 Accession Number

Cd. No.	Glossary	Term
9 10	15 16	21 80
9 0	0.6	SHRINKAGE
9 0	0.6	AESTHETIC QUALITIES
9 0	0.6	MESH FABRICS
9 0	0.6	NETS
9 0	0.6	LOOP STRENGTH
9 0	0.6	SEAM STRENGTH
9 0	0.6	TOUGHNESS
9 0	0.6	DIMENSIONAL STABILITY
9 0	0.6	STRETCH PROPERTIES
9 0	0.6	TENSILE TESTS
9 0	0.6	GAGE TESTS
9 0	0.6	VARIABILITY
9 0	0.6	[illegible] LIMIT
9 0	0.6	EVALUATION
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		
9 0		

GLOSSARIES

01 Originating Activity
02 Sponsor
03 Contract Number
04 Publication Source
05 Personal Author
06 Institute of Textile Technology
07 Engineers Joint Council
08 Defense Documentation Center
09 American Inst. of Chem. Eng.

Figure 4 Document Subject Indexing Keypunch Format

Modern Textiles 47/9 Sep 66

DURABLE PRESS

The Role of Polyesters

by Wells Denyes
Senior Textile Engineer
Textile Development Division
Tennessee Eastman Co.

WITHOUT QUESTION, the most exciting new phrase in the entire textile industry right now is "Durable Press." From the research chemist to the housewife, textile products today cannot be discussed without mentioning the new term. Researchers are already seeking new resins, new fibers, new fabrics, and new applications for durable press. Housewives are busy reseeking the family wardrobes and furnishings to take full advantage of the many wonderful properties of durable press.

Figure 7 Cut Title Page

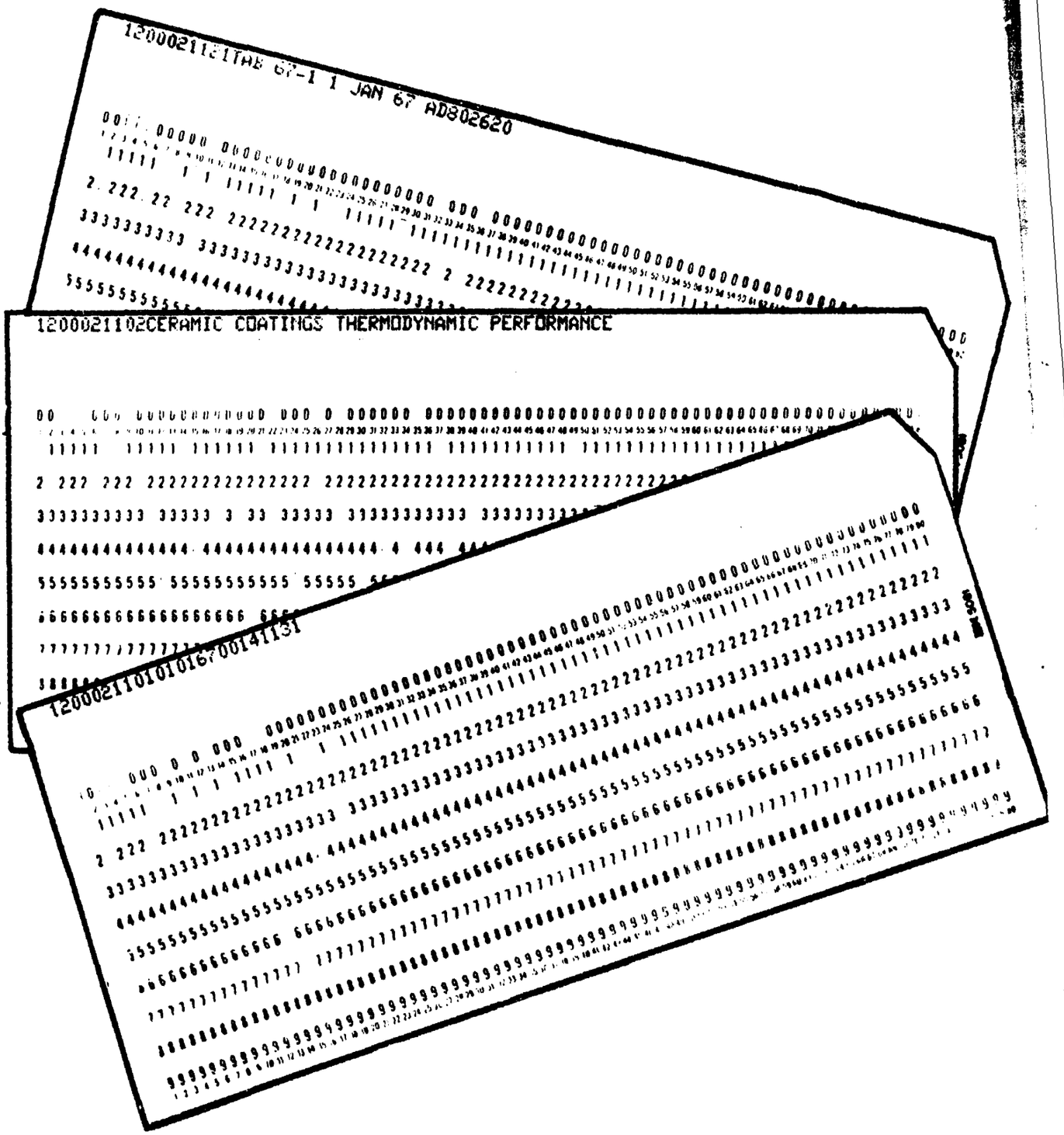


Figure 8 Title and Journal Citation Punched Cards

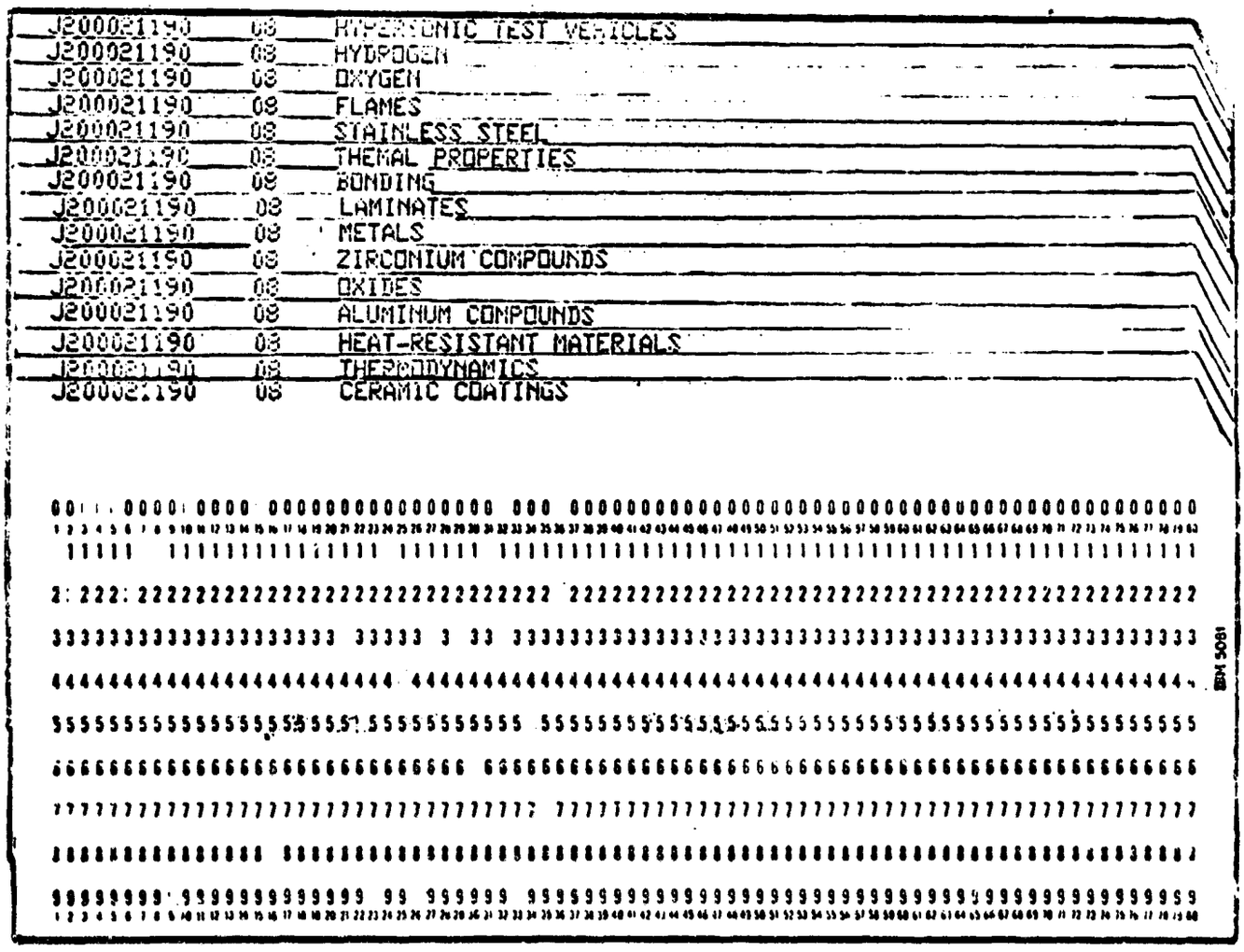


Figure 9 Document Term Punched Cards

☆ Textile T[illegible]nology Terms: An Information Retrieval [illegible]esaurus, first edition, second printing, February 1966, Textile Information Center, Institute of Textile Technology, Charlottesville, Virginia

☆ Thesaurus of Engineering Terms, first edition, May 1964, Engineers Joint Council, New York, New York

☆ Thesaurus of ASTIA Descriptors, second edition, December 1962, Armed Services Technical Information Agency, Arlington, Virginia.

In the subject area of textile technology, a thesaurus comparable to that published by ITT was located at M. I. T. This was the Thesaurus of Textile Engineering Terms, first edition, February 1966, by Stanley Backer, et al, M. I. T. Department of Mechanical Engineering, Fibres and Polymers Division. The Thesaurus of Astia Descriptors was supplemented by the Astia Chemical Thesaurus, first edition, December 1962, the DDC Supplement to the Thesaurus of Astia Descriptors, second edition, April 1964, the DDC Thesaurus Code Manual, second edition, August 1963, and the DDC Descriptor Frequencies, May 1966. In addition, terms were also used as they appeared in the DDC Technical Abstract Bulletin appended to abstracts selected for dissemination. A publication received too late for use in the pilot program was Glossary of Plastics Terms: A Consensus, December 1966, Plastec, Plastics Technical Evaluation Center, Dover, New Jersey. This is an excellent glossary of the commonly used technical terms in the fields of plastics, adhesives, and elastomers.

As could be expected, there was a considerable overlap among the terms used in the three thesauri, the greatest between the Thesaurus of Astia Descriptors and the Thesaurus of Engineering Terms. The former was most distinctive in the areas of military applications, whereas the latter was most distinctive in the areas of plastics, polymers and elastomers. Each document (journal article) was indexed primarily from only one of the three thesauri. The choice of thesaurus was guided by the following table:

(1) Thesaurus of Engineering Terms used for plastics, physical properties of materials, high-vacuum techniques and equipment,

physical measurements, analytical instrumentation and refrigeration.

(2) Thesaurus of Astia Descriptors for chemical, biological, and radiological warfare, chemistry (organic and inorganic), protective clothing, and space technology.

(3) Textile Technology Terms used for synthetic fibers, textiles, dyeing, detergents, fabric properties, fiber properties, and clothing manufacture.

There were several difficulties encountered in utilizing three thesauri for indexing both documents and interest profiles. The most common was that a document could logically be indexed by more than one thesaurus so that personal preference of the indexer played some part in determining which thesaurus would be used. This caused a loss in potential selections due to the fact that a particular interest profile may not have contained some of the terms from the thesaurus used to index a document relevant for that individual. A second difficulty was the lack of a valid term or terms to describe a particular document or interest. This situation most frequently related to articles dealing with research in organic chemistry. Understandably, many of the specific chemical names were absent from all three thesauri.

To prevent the loss of valid selections, each profile was supplemented by terms from all three thesauri. This caused an overexpansion of the user profile, but yielded a higher recall ratio. A much needed exercise, but one outside the scope of this contract, is that of concatenating all of the thesauri pertinent to NLABS endeavors. This project would involve the keypunching of approximately 50,000 terms and manual sifting to eliminate duplicate terms and synonyms, and would have to be undertaken prior to implementation of the full-scale SDI system. Although the use of multiple glossaries is feasible during a pilot test, involving a limited document sample and population of users, it would be unworkable in full-scale implementation.

2. Glossary Organization and Coding

Each glossary was assigned a unique two-digit number. The glossary number combined with the alphabetic spelling of the term constituted the unit for comparison between document terms and interest profile

terms. Even though a term may have been spelled identically in two different glossaries, a match between a profile term and document term would not occur unless the glossary numbers were identical. This rigid association of term and glossary was maintained in order to trace the effect of the choice of glossary on the selection process. In addition to a glossary number, each term could be assigned a unique four-digit number to identify it within the glossary. The term number combined with the glossary number could be used in the matching process in lieu of matching the alphabetic term itself.

Numbers could be assigned so that a numeric sort could accomplish an alphabetic sorting of the terms within a specific thesaurus. This practice would ultimately facilitate the conversion of the match and selection process to an inverted file so that matching time by computer could be substantially reduced (See Figure 10).

E. USER INTEREST PROFILE CONSTRUCTION

Construction of the user profile is the most critical and the most difficult aspect of selective dissemination of information. The task of creating interest profiles can be much simplified by a coherently organized source of technical terms. The fact that a single coherent indexing tool was not available during the conduct of the pilot test understandably overcomplicated the task of creating interest profiles. However, the techniques used to circumvent this difficulty allowed for the creation of valid profiles and for the testing of several different profiling techniques.

As pointed out in Section IV, B, computer programs were written to implement the concurrent use of two different match strategies. Both strategies exemplify a multidimensional retrieval approach. Dimensions consisted of term sets exemplifying different points of view or different aspects of the same interest. For instance, an interest may be expressed as "diffusion of oxygen, nitrogen, and carbon dioxide through polyethylene and other polymer films." The different aspects of this interest are: (1) the process (diffusion); (2) the physical state (gas) and chemical composition (oxygen, etc.) of the diffusing substance; (3) the composition of the barrier (polyethylene); and (4) the physical form of the barrier (film). Any one dimension might prove to be the most limiting factor in retrieval. That is, if the document collection contains a minimal number of articles on gasses, then this facet of the interest profile would be the most limiting. The ability to take advantage of the selectivity of one dimension of a multidimensional interest enhances both relevance and recall. It was for this

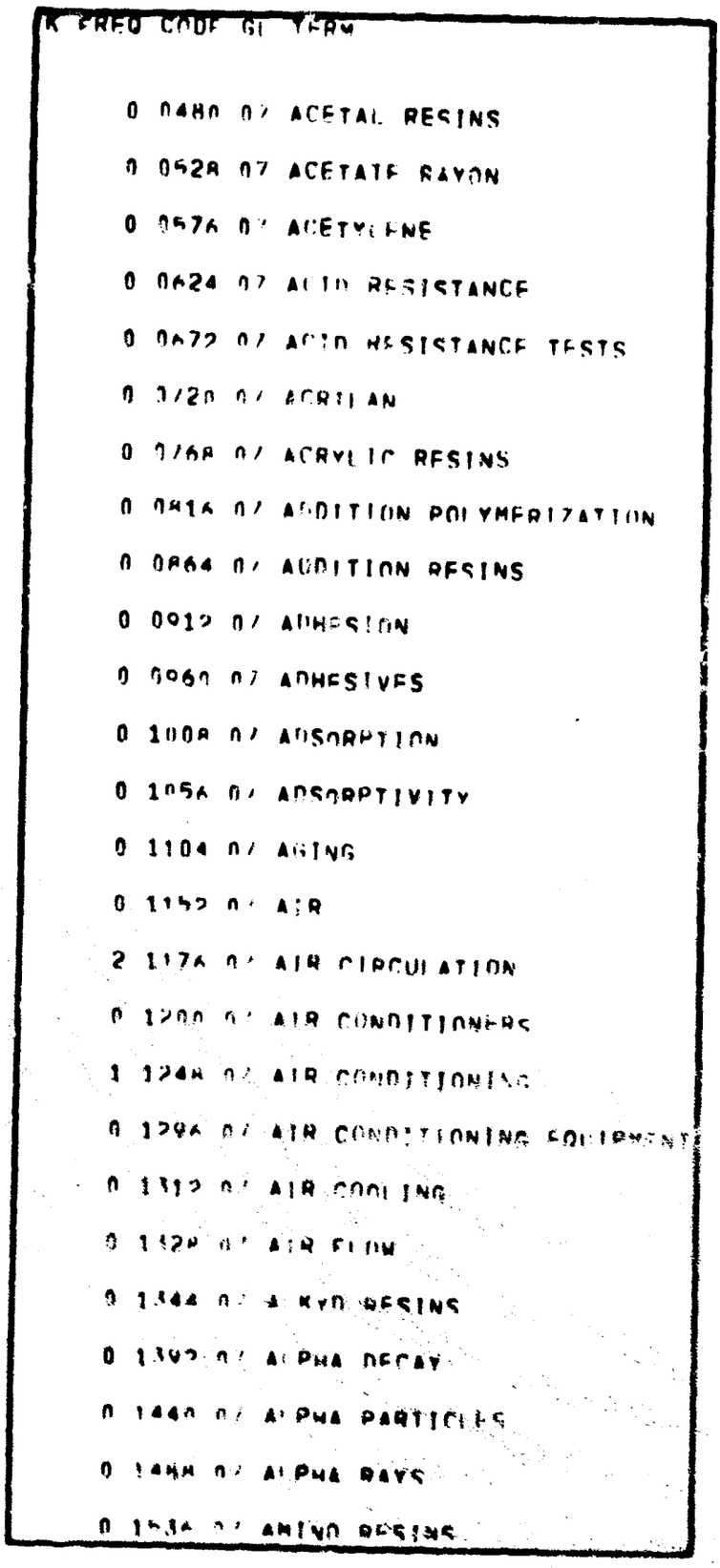


Figure 10 Glossary

reason that the Boolean and weighted-term strategies were selected for testing in this pilot program.

1. Structure of an Interest Profile

Each interest profile consisted of one or more sets of terms descriptive of each aspect of a multidimensional interest. In testing the weighted-term strategy, all terms in an interest profile were carried in the same set so that essentially the profile was single dimensional. Dimensions were simulated, however, by giving certain terms more weight than others. In testing the Boolean strategy, all terms were given the highest possible weight, but like terms were grouped together. The match strategy required that at least one term from each group match a term in the document description. The two methods were combined for some profiles, so that selection was based on both term grouping and term weights.

In the first dissemination cycle, all profiles were constructed according to the weight strategy with every term carrying the highest possible weight. Thus, a selection would be assured if any one profile term matched a document description term. In the next dissemination cycle, some of the profiles contained multiple sets of terms and some contained only one set of terms. All weights were set at .5 so that selection would occur only if two or more profile terms matched document terms. Profiles were modified from the first cycle by addition of terms included in documents selected in cycle I. Terms were also added from documents processed in cycle I, but not automatically selected. (Participants were allowed to scan a list of the titles of documents processed in cycle I and to select the titles that appeared to be interesting which were not selected automatically for them.) For cycle III, profiles were extensively reworked to eliminate terms which were suspected of causing irrelevant selections and by combining terms from different glossaries into the same profile. The latter move was made to reduce processing time and had no effect on selection. In cycle III, the profile structure remained basically the same, except that some profiles were changed from the Boolean to the weighted strategy. In cycle IV, profiles were again reworked on the basis of response to selections from cycle III. In this cycle, weights were varied to determine the effect on selection.

2. Preliminary Screening of Terminology

Derivation of the initial interest profile was the most difficult phase of profiling. Because of the use of several thesauri and a

limited number of copies of each, a decision was made to present each participant with a subset of terms which were thought to reflect his interests based on his own statement. These sets of terms were assembled into a folder for each participant. At a group meeting, participants were asked to select the terms from the list in their folder and to record them on a preprinted profiling form. The participants were also allowed to supplement the terms available to them by scanning the three thesauri. This process was necessary for more than half of the participants, since the terms provided were not sufficient. However, all participants expressed a favorable opinion of being provided with an abbreviated version of the thesauri that was tailored to their expressed interests. Even though more than half of the participants selected additional terms direct from the thesauri, most of the terms which they recorded were obtained from the preprinted lists.

Profiling was accomplished in three, two-hour profiling sessions. Members in the first group were allowed to structure their profiles according to the Boolean strategy. Members in the second group used a weighted strategy. Third-group participants were allowed to use a combination of the two methods. For the initial SDI cycles, however, these structures were temporarily suspended until further experience could be obtained on the percentage of matches that would occur with the document sample.

3. Preparation of Interest Profiles

Initial interest profiles were keypunched from worksheets which contained four preprinted columns (See Figure 11). Columns were designated for profile number, term set number, assigned weight, and term with glossary number. Initially, profiles were punched with varying weights and levels as written, but after the decision had been made to perform a test cycle using single-aspect profiles and maximum assigned weights, the punched cards were duplicated and a standard set number equal to one weight of .999 were gang punched into the entire profile deck. In subsequent runs, levels were restored to the profiles and weights were varied from the minimum of .999. Profiles in cycle II were expanded over 200% from those used in cycle I. In cycle III, profiles were reduced from cycle II by 30%. Cycle IV profiles required very few changes over cycle III profiles. The substantial additions made over cycle I profiles necessitated keypunching a significant number of additional cards. Changes on cycle II profiles, however, required mainly the elimination of cards. Cycle IV changes required the addition of a relatively small number of terms.

F. DISSEMINATION OF INFORMATION

SDI notices were prepared by conventional office duplicating techniques and collated for distribution on EAM card processing equipment.

SUBSCRIBER PROFILE WORKSHEET

1 8

1,3 Subscriber

Header		
9 10 11	30 31	79 80
Card	Name	Address 1 2
0,0		

1 8

J,3 Subscriber

Profiles					
9 10 11 12 13 14 15 16 21					80
Prof*	L	Weight	GL	TERM	
0,1					
		—	—	—	

1 Availability

1 For Info Only
2 Will Obtain
3 For Ref Only
4 For Circulation
5 Personal Copy

2 Notice Format

1 Title
2 Terms
3 Abstract

3 Glossary

01 Originating Activity
02 Sponsor
03 Contract No.
04 Publication Source
05 Personal Author

* Profiles 01 thru 99

Figure 11 Profile Worksheet

1. Preparation of SDI Notices

The title page of each journal article was copied onto 8 1/2 x 14" paper on a Xerox #914 Office Copier (See Figure 5) which allows bound volumes to be copied without destroying the binding. The journal page was positioned to allow a free space at the bottom of the copy approximately three inches high. A preprinted keypunch form and a typed strip of paper designating the journal title, volume, number, and date were placed in this area. The keypunched form designated three card types and the information to be keypunched in each card field. The complete Xerox copy thus provided all the descriptive cataloging information, including the title, journal citation, authors, author affiliation, contract support, and translation information. During the pilot test, only title and journal source were recorded in machine-readable form. While each title page was being keypunched, subject indexing was performed on the original journal article, and subject terms were recorded on a second preprinted keypunch form (See Figure 6).

On return of the document title page copy, punched cards were proofread against the title page and corrected. At the conclusion of the correction process, the information to appear on an SDI notice was clipped from the title page and stored in a numbered envelope (See Figure 7). This information included the title and other descriptive cataloging information and the abstract, introduction, or some portion of the article itself. These clippings were stored in numbered envelopes filed in numerical sequence according to an assigned accession number stamped on the article title page at the time of selection.

Terms keypunched from the document and subscriber profile worksheets are input to the computerized match function. The computer program matches document terms on one magnetic tape against profile terms on another.

The output of the computer matching process is a punched card deck containing two types of cards. One type contains the document number and the number of times that document was selected for dissemination (See Figure 12). The other card contains a document number and a subscriber and profile number (See Figure 13). After the two card types are separated by sorting, one card deck is sorted into numerical order by the number of copies of each document required. The sorted deck is listed to provide directions for the notice duplication process (See Figure 14).

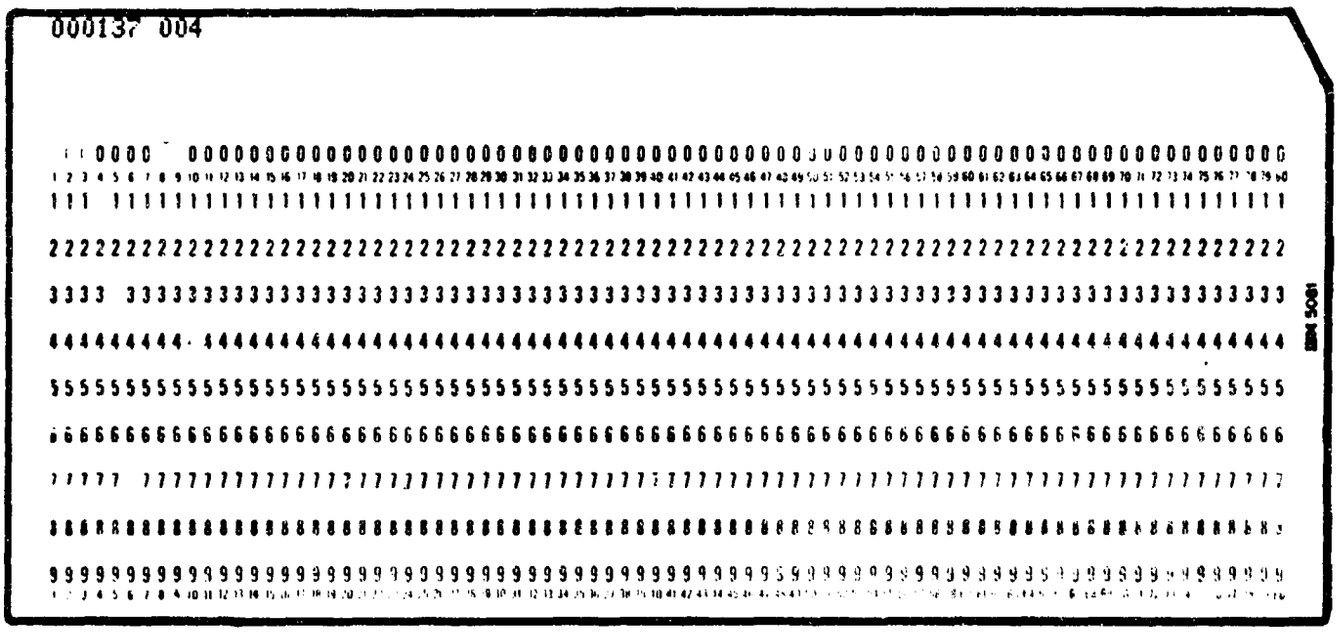


Figure 12 Copy Card

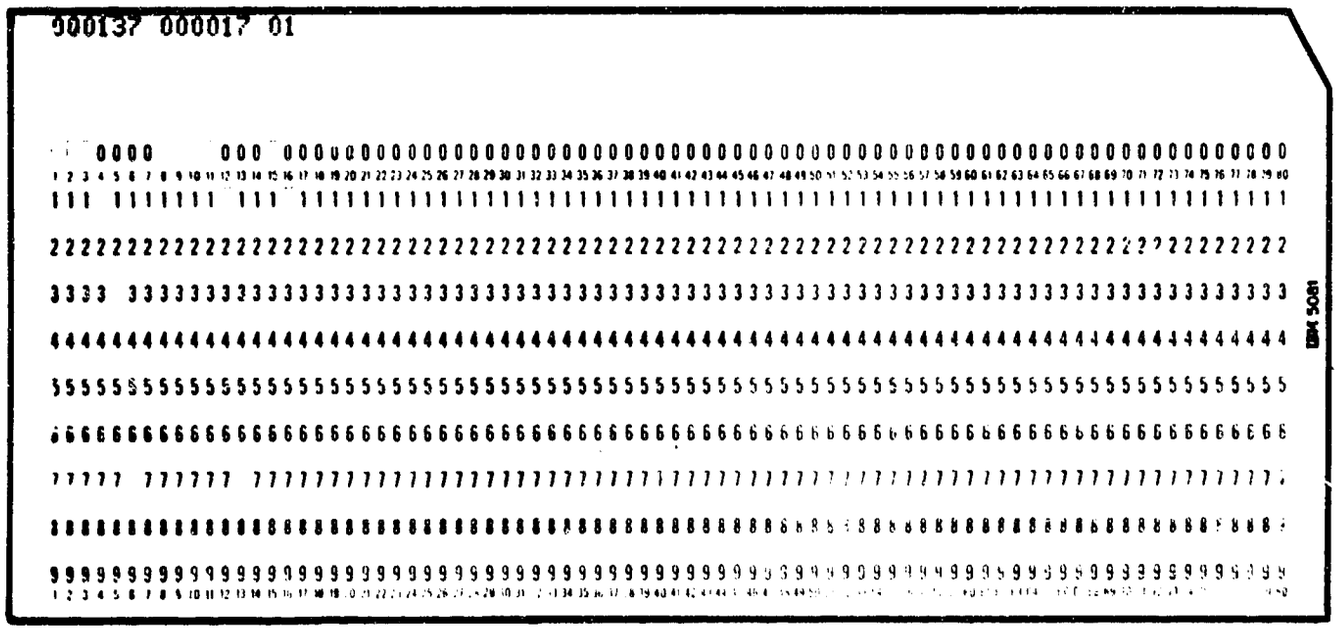


Figure 13 Distribution Card

```
000848 004
000892 004
000897 004
000899 004
000906 004
000912 004
000913 004
000935 004
000936 004
000958 004
000986 004
001003 004
001006 004
000795 005
000812 005
000817 005
000818 005
000825 005
000836 005
000837 005
000845 005
000853 005
000904 005
000910 005
000964 005
000994 005
001004 005
001005 005
```

Figure 14 Copy Card Listing

During pilot operation, the maximum number of copies required of any one SDI notice was fourteen out of a possible twenty five. Since the mean number of copies required was three, reproduction by printing was uneconomical, since press set-up time and the master preparation would cost more than conventional office copying. For this reason, all SDI notice copies were prepared on the Xerox 914 using prescored four-up IBM card stock (See Figure 15). This stock is a recently developed item available from Essex Systems Incorporated in quantities of 1000 sheets (or 4000 cards) per case.

The printed list containing the number of copies required of each document (See Figure 14) is used as a basis for assembling the clipped sections of journal article title pages into master sheets requiring the same number of duplicate copies. First, all the clippings from which one copy is to be made are pulled from the numerically stored envelopes; then, all clippings requiring two copies are pulled, etc. The content of each envelope is fixed to a format sheet blocked out to correspond with the four IBM card-sized image areas of the Essex card stock (See Figure 16). All clippings requiring one copy are affixed to as many sheets as necessary. The same process is required for all of those clippings requiring two copies, three copies, etc.

Each format sheet containing up to four card images is placed face down on the Xerox 914 copy glass, and the Essex card stock is loaded into the paper stock feeder. In some instances, two sides of a card are required to carry the information to be disseminated. When this is necessary, the flip-side image is properly positioned on the reverse side of the format sheet. After duplication of the front side, the card stock is turned over, reloaded into the feeder, and the flip-side image is copied.

At the conclusion of the duplicating process, the card stock is bursted (See Figures 17 and 18) and the required notice copies are assembled behind the punched and interpreted copy card generated by the computer matching. The assembled card deck thus contains a header card punched and interpreted for document number and number of copies, followed by the appropriate number of imprinted unpunched notice cards. This card deck is then intersperse gangpunched. Punching transfers the document number in the header card into each one of the notice cards behind it. A first sort removes the header card, the remaining card deck is then sorted in numerical order by document number (See Figure 19).

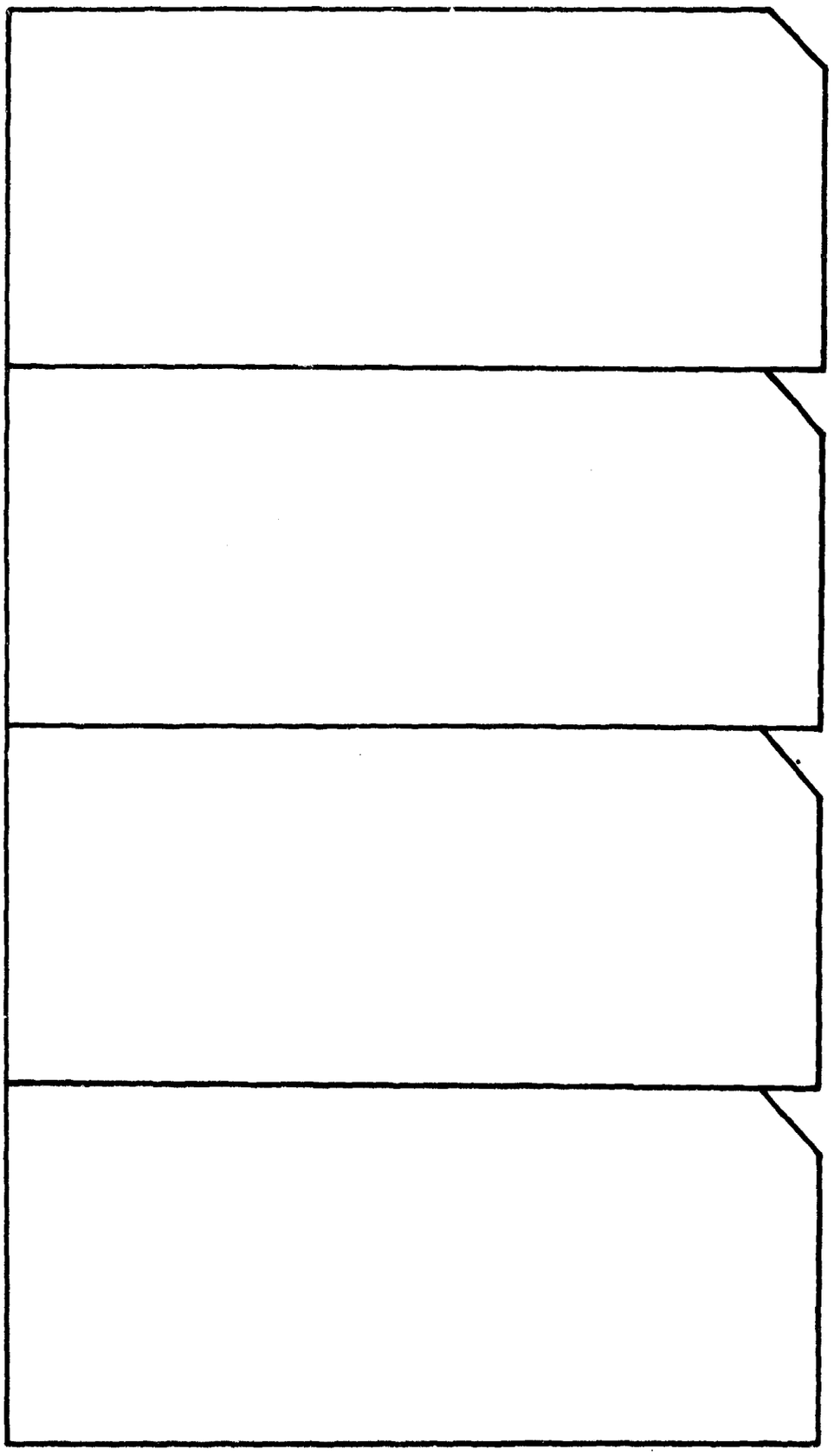


Figure 15 Four-up Pre-scored IBM Card Stock

J of Colloid & Interface Sci 23/1 Jan 67

(0) 0823

Adsorption of a Nonionic Surfactant by Cotton[1]

HANS SCHOTT

ABSTRACT

The adsorption and desorption of polyoxyethylated 1-dodecanol containing an average of 11 ethylene oxide (EO) units on cotton was studied over a concentration range of two decades in distilled water at 25 C.

The method consisted in equilibrating cotton fabric or chopped fiber with surfactant solutions and determining the equilibrium surfactant concentration by surface tension or by the cobalt blue colorimetric analysis. Surface tension is more affected by surfactant molecules of few EO units, whereas the colorimetric method emphasizes the more highly polyoxyethylated surfactant molecules; the observed discrepancy between the two methods (10%, 30%) was caused by preferential adsorption of surfactant of shorter EO chain length onto cotton.

The adsorption was almost completely reversible. The adsorption isotherm leveled off at 35 mg. surfactant/100 g. cotton near the critical micelle concentration. In view of the specific surface area of cotton, this level of adsorbed surfactant is of the order of the amount required to cover the cotton completely with a close packed monolayer of surfactant molecules lying flat against the substrate.

J of Chem Physics 46/2 15 Jan 67

Surface Dye Sensitization. II. Sulfur

(0) 0824

S. W. ING, JR., AND Y. S. CHIANG

Research Laboratories, Xerox Corporation, Webster, New York

(Received 28 July 1966)

I. INTRODUCTION

SULFUR, being a wide-bandgap material, has very little photoelectric response to visible light. Spear and Adams[1] have studied the electronic structures in orthorhombic sulfur crystals. They conclude, tentatively, that there is a valence band for hole conduction; however, the electrons move about by a hopping process. The intrinsic bandgap energy for hole–electron generation is reported[1] to lie in the range from 3.65 to 4.30 eV.

We have observed that certain dyes applied to the surface of the crystalline S (orthorhombic) can extend its photoelectric response to the visible region. Sulfur is the largest-bandgap inorganic material we have seen which exhibits this dye-sensitized response.

Figure 16 Paste-up of Card Images (Four-up)

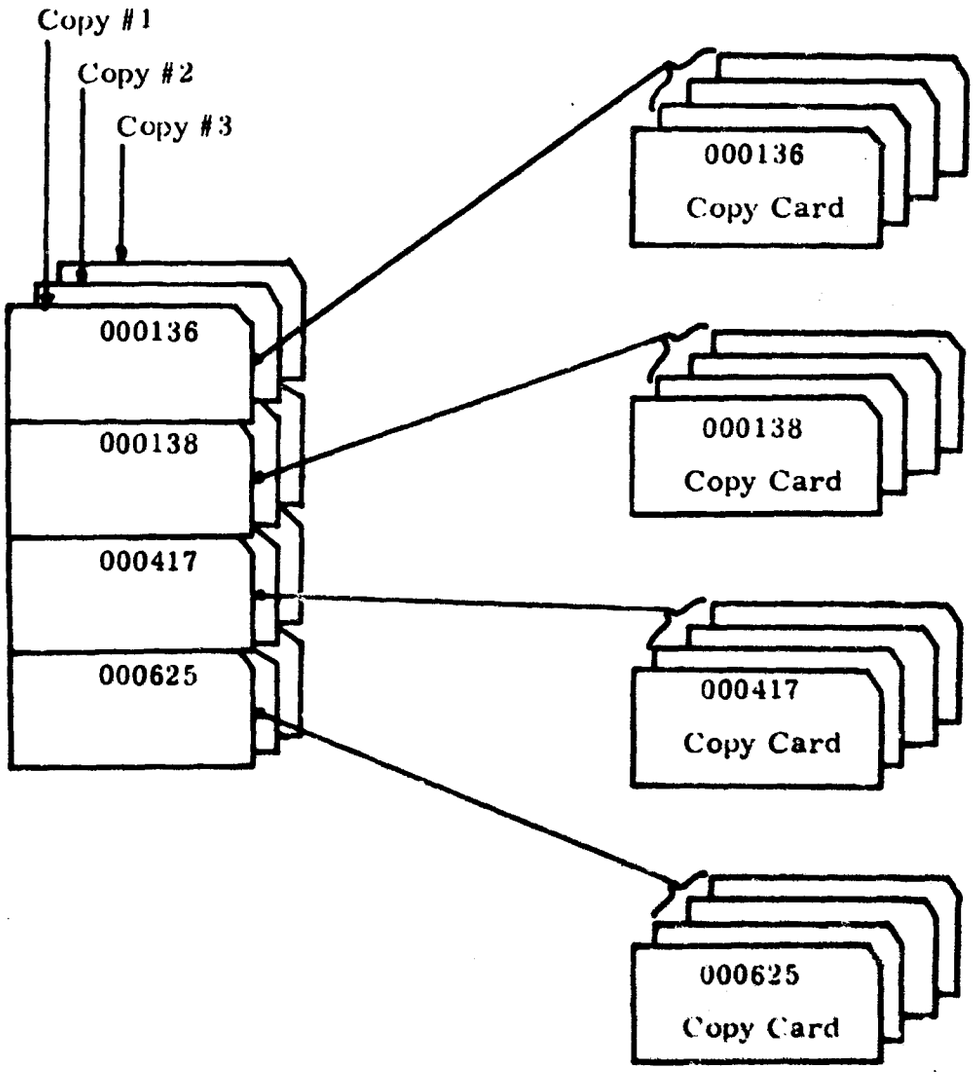


Figure 17 Bursting of Cardstock

16 000012 04

TAB 67-3 1 Feb 67

AD-644 118 Fld. 6/17, 6/19
CFS. 1 Prices: HC $3.00 MF $0.65
CIVIL AEROMEDICAL INST OKLAHOMA CITY OKLA
EVALUATION OF THE PHYSIOLOGICAL PROTECTIVE EFFICIENCY OF A NEW PROTOTYPE DISPOSABLE PASSENGER OXYGEN MASK.
by Ernest B McLadden Apr 66 27p
AM 66 7

Unclassified report

Descriptors: Oxygen masks, Efficiency, Breathing apparatus, Protective clothing, Responses, Respiration, Altitude chambers, Body temperature, Oximeters, Exercise, Experimental design, Aviation medicine

A prototype of a new design disposable passenger mask applicable for emergency use in jet transports at altitudes to 40,000 feet was evaluated. Six subjects instrumented to obtain a variety of physiological information were exposed to a chamber flight profile designed around the National Aerospace Standard 1179. The two alternative methods of determining mask performance suggested in this document and based on gas analysis and blood oxygen saturation were used simultaneously in this study. In order to stimulate the respiration to the 30 liters/minute volume levels specified in applicable regulations, the subject exercised on a bicycle ergometer through the chamber flight up to and including 40,000 feet. Air breathing baselines were established at 10,000 and 14,000 feet with the subject resting. The 14,000-foot baseline determination was repeated at 14,000 feet exercising at the same work load level as used in the high-altitude tests. The mask maintained all subjects in a satisfactory physiological condition at all altitudes up to and including 40,000 feet for the duration of exposure used in these tests. (Author)

44 000000 C1 Polymer Engrg & Sci 7/1 Jan 67

20 0094

An Infrared Spectroscopic Investigation of Photodegraded Styrene Copolymer Films

Raymond B. Seymour, Hing-Shya E. Tsang, Doug Warren,* *Chemistry Department, University of Houston*

The degradation of organic polymers in the presence of ultraviolet light is technically and economically important. Considerable emphasis has been placed on the development of test procedures and the investigation of the mechanism of photodegradation.

Committee D-20 of the American Society for Testing Materials has established a subcommittee to evaluate the effect of light on permanence properties of plastics. Recommended practices have been outlined for outdoor (1) and accelerated weathering tests (2). In one test procedure, the intensity of natural sunlight is increased by the use of mirrors (3).

Figure 18 Single Card Notice

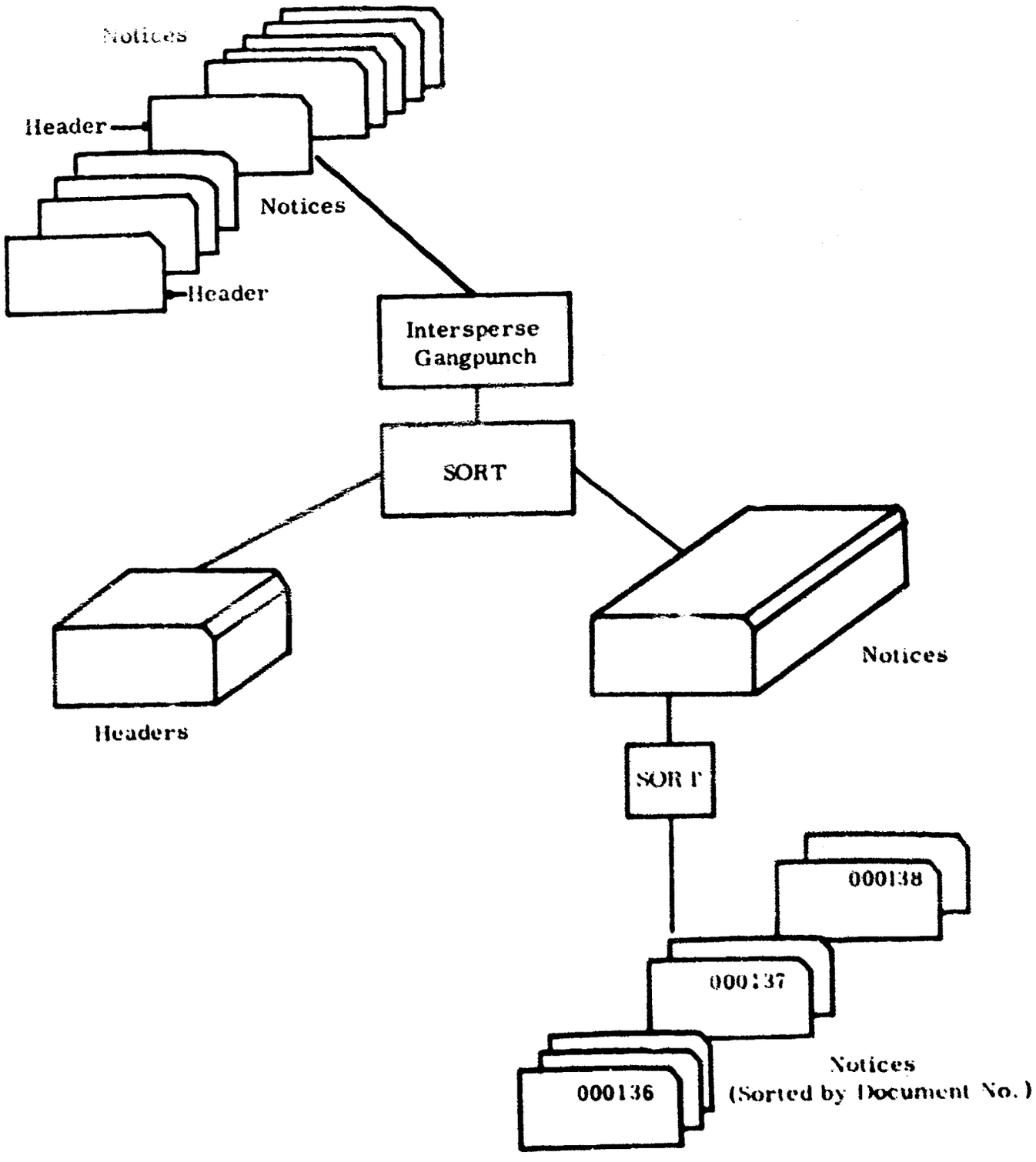


Figure 19 Notice Card Sort

2. Collation and Delivery of Notices

The distribution cards punched with a document number and subscriber and profile numbers are also sorted into numerical order by document number (See Figure 20). These cards are duplicated into a preprinted response card (See Figure 21). The response cards are then collated with the notice cards so that response cards alternate with notice cards in the resulting merged deck. This deck is then intersperse gangpunched so that the subscriber and profile numbers are transferred to the notice card behind each response card. Punching allows the deck to be sorted into numerical order by profile number so that all notices and response cards to be sent to one participant fall in the same sorting pocket.

During the pilot operation, packets of notice and response cards were addressed by hand. In full-scale operation, this function can be accomplished by merging, by means of EAM equipment, an address header card, prepunched and interpreted and containing a distinctive colored stripe to act as a separator between the packets of notice and response cards.

To expedite the evaluation of the notice cards and return of the response cards, each packet of response and notice cards was hand-delivered to each participant in the pilot program.

G. RESPONSE FROM PARTICIPANTS

Participant responses were obtained orally at the time of notice delivery and by means of the preprinted response card associated with each notice delivered.

1. The Response Form

The response form (See Figure 21) was divided into four quarters of approximately equal size. The first quarter contained document number, subscriber number, and profile number interpreted at the upper left hand edge of the card. The second quarter contained marking positions for six comments: of interest, notify when available; of interest, document not wanted; of interest, have seen copy; of no interest; refer notice to Mr. _____; and, modify my profile--changes below. The third quadrant contained space for profile modification, and the fourth quadrant contained instructions to the user.

Since the response and the notice cards were hand delivered, most of the comments pertaining to profile changes were received in conversation, rather than in the response cards. Thus the efficiency of this block could not be thoroughly tested during the pilot test.

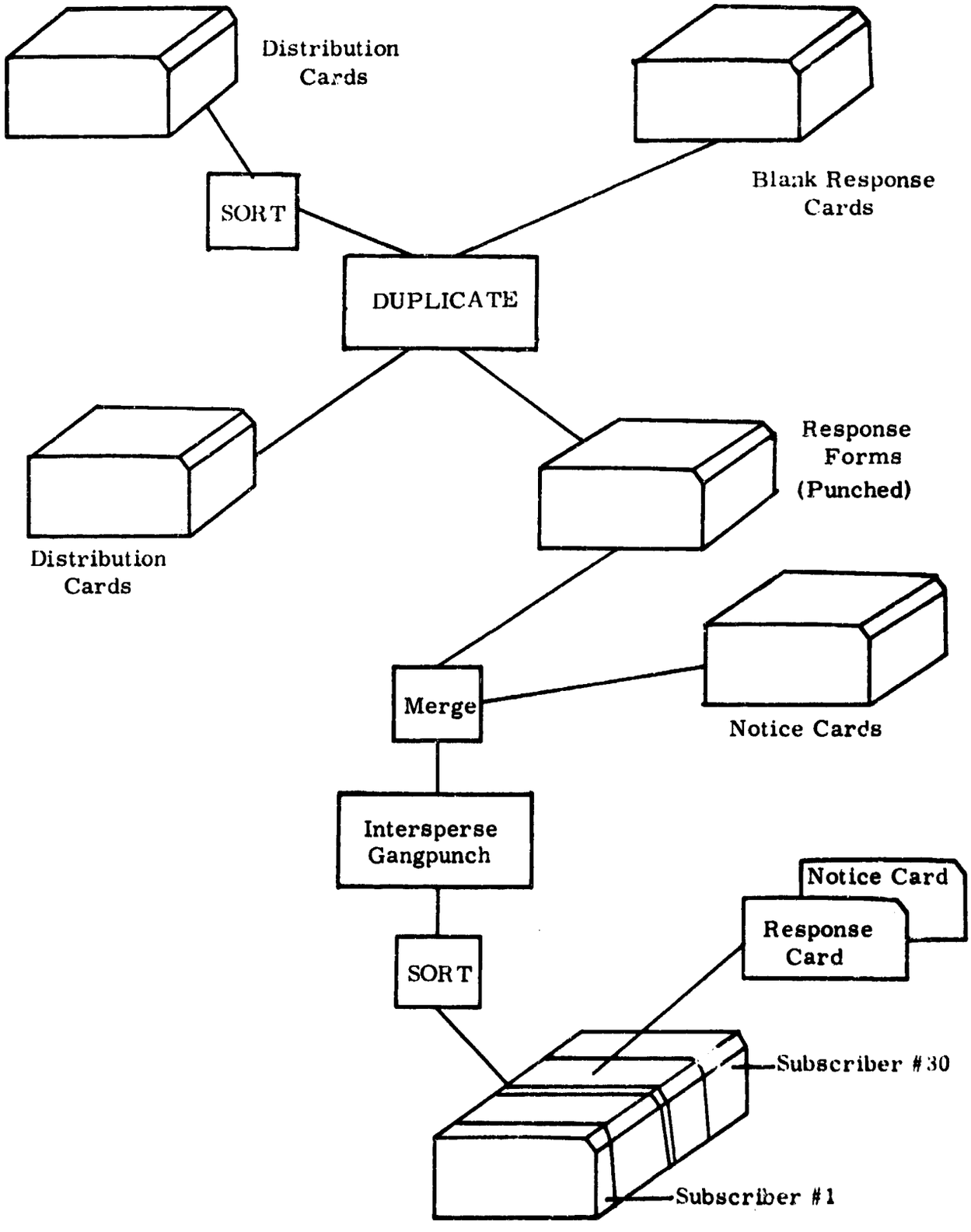


Figure 26 Distribution Card Sort & Collation With Notice Cards

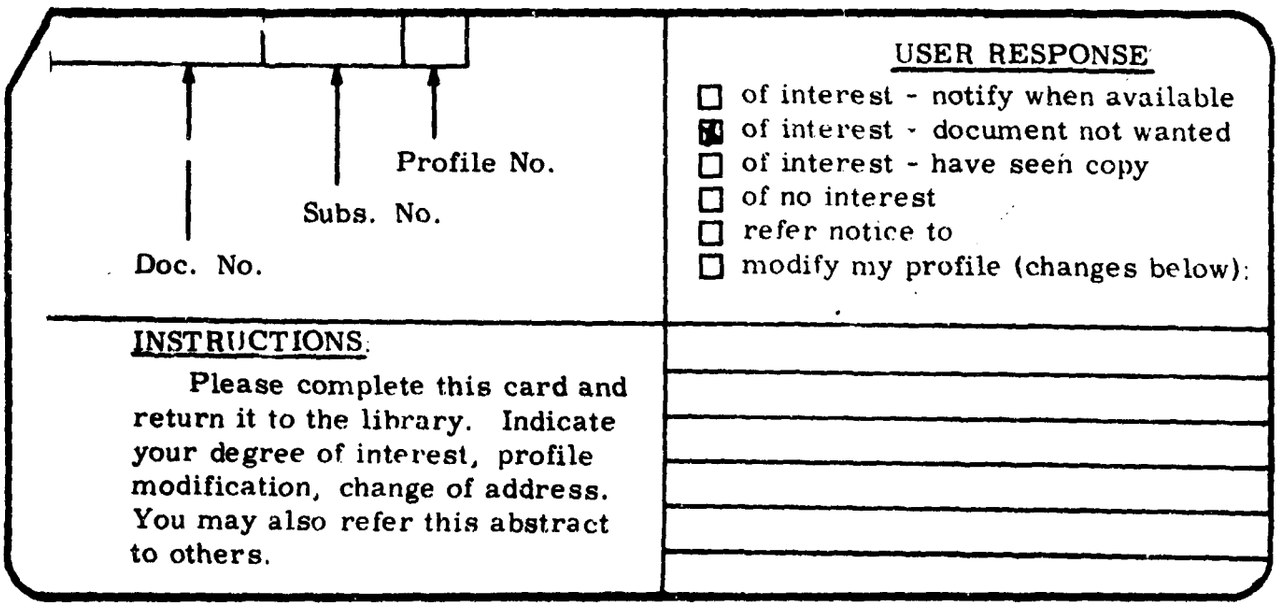


Figure 21 Response Card

Since the response and the notice cards were hand delivered, most of the comments pertaining to profile changes were received in conversation, rather than in the response cards. Thus the efficiency of this block could not be throughly tested during the pilot test.

A few participants showed a marked interest in referring notices to fellow workers who were not officially participating in the SDI program. Relatively few referrals were made, however, and these were made by few participants. Comments most frequently checked were: of interest, notify when available; of interest, document not wanted; and of no interest.

Participants recorded the titles of the journals which they regularly read prior to the first SDI cycle. Although many of the journals processed during the the pilot test were listed by the participants, the comment "of interest -- have seen copy" was checked very infrequently. It can only be conjectured at this time that the information was being delivered through the SDI channels at a more rapid rate than would normally be achieved during regular reading of the journals themselves. It cannot be assumed that the participants would not have seen all of the articles which were of interest to them as marked on the response cards. The fact that these journals were out of circulation during the processing may have influenced this factor also.

2. Analysis of the Responses Returned

Primarily, responses were analyzed to determine the effect of varying the search strategy. Tabulations of response versus strategy were analyzed to discover whether one or the other or a combination of Boolean and the weighted search strategies were more appropriate and yielded a higher relevancy rate.

In addition, responses were analyzed to discover what correlation existed between relevance and (1) type of document, (2) number of profile terms, (3) number of levels per profile, (4) notice format, (5) job type, and other factors.

3. Adjustment of Interest Profiles

Profiles were adjusted after each SDI cycle. Following the first cycle, terms used to index documents selected and marked of interest by a participant were added to his profile. In addition, each participant

scanned a list of titles processed during the first cycle and marked those which he thought might be of interest. Terms used to index these articles were also added to his profile.

After the second SDI cycle, terms which were thought to cause a selection of irrelevant documents were removed from the participant's profile. In general, the profiles were condensed in size and the total number of profiles per participant was condensed in order to speed up processing time. An attempt was also made to select the most limited factor exhibited in each profile so that searching time would not be used needlessly in matching frequently occurring terms when selection would be limited ultimately by an infrequently occurring term. The grouping of terms into sets was also rearranged so that all terms in a set would have a similar relationship to each other.

After the third dissemination cycle, profiles were refined by minor adjustments in the grouping of terms and the assignment of weights.

V ANALYSIS OF PILOT TEST RESULTS

Judging from the results of the pilot test conducted at Army Natick Laboratories, it is possible to predict that a similar SDI system can be installed to support the technical personnel at other Army research installations. Since the requirements met at the NLABS installation were the most stringent in terms of breadth and depth of subject coverage, an even higher success rate in terms of average relevancy of the information disseminated can be predicted for other installations. In the following analysis, the SDI pilot test is evaluated in terms of overall effectiveness, operating costs, and effort necessary to implement a similar SDI system at other Army installations. In addition, the statistics on the four SDI production runs are analyzed in terms of the various parameters that affect cost and relevancy.

A. EVALUATION OF EFFECTIVENESS OF SDI

1. Effect of Limitations on the Pilot Test

The most obvious limitation on the pilot test was the selection of a limited number of participants. Although this limitation precluded the incorporation of the complete range of interests existing at NLABS, this limitation did not impair the validity of the results obtained. Interest profiles of the twenty-five participants selected were sufficiently representative of the total range of interest so that results can legitimately be extrapolated to those obtainable under full-scale operation. The depth of interest existing in each profile was representative of the variations in depth that might be encountered in a full-scale operation. This range in depth of interest varied from one participant, whose job was the writing of specifications, to another participant whose job was the synthesis of new acetylenic organometallic compounds of Group IV and V metals. These extremes in depth of interest can be detected in the statistics on profile structure presented in Appendix A. The very broad interest is represented by a profile containing a large number of terms, whereas the very specific interests of the organic chemist are represented by less than fifteen terms.

Temporal and monetary limitations on the pilot test had more significant effect on the results obtained than did the limitations on the number of participants selected. Implicit in the conduct of a pilot test is the stringent limitation on the time available. In a full-scale implementation it will be possible to perform a wider variety of tests and to obtain more detailed explanations of the phenomenon which occurred during the pilot test. The substantiation of some of the hypotheses presented in this analysis must await further exploration. Results of the pilot test indicate that it may be possible to create a standard technique for evaluating all SDI

systems on a normalized basis. Some of these analytical techniques are included in this section of the report to indicate a topic for additional research.

The most limiting factor affecting the outcome of the SDI pilot test was the lack of a completely integrated subject vocabulary. The use of externally-produced thesauri, dictionaries, and glossaries provides only a partial solution to the larger vocabulary problem. The results of the pilot test indicate that available term lists provide terms which are too narrow in scope for use in selective dissemination. The term lists used were structured in a hierarchical array at a depth too low for practical full-scale use.

A specific example will clarify this problem. A typical interest in the pilot test was "high-temperature properties of materials for protection against thermal radiation." Glossary terms which apply to this interest are the following: high temperature tests; pyrolysis; ablation; oxidation; thermal degradation; thermal stability; melting points; thermal analysis; high temperature; thermal properties; specific heat; calorimetry; infrared radiation; nuclear blast effects; thermal energy; plastics; polymers; fibers (synthetic); fibers (natural); etc. Since the indexer has such a wide variety of terms from which he can select a subset to describe a particular document, each profile must contain a comparable number of terms to increase the probability that a profile term will match a document term. However, this strategy vastly increases the number of irrelevant matches since the total number of term combinations that might cause selections is proportional to a factorial function of the number of terms in the subscriber's profile. What is needed to increase the efficiency of the matching process and the relevancy of information distributed is a complete and thorough analysis and structuring of a vocabulary according to the specific requirements of the users of the system. This endeavor is properly the subject of a separate high-level endeavor utilizing specialists knowledgeable in each of the areas of interest to the participants.

The creation of a comprehensive and coherent vocabulary would not only increase the efficiency of the selection process and the relevancy of disseminated information, but would also speed up the document indexing and subscriber profiling processes. If terms were imbedded in a complete hierarchy, the indexer could select terms down to the exact level required. The length of time required for indexing is roughly proportional to the number of terms that must be selected to represent the subject content of a document. Using pre-existing vocabularies arrayed in a two-level hierarchy required the indexer to select an inordinately large number of specific terms to represent a general topic. Correspondingly,

the number of terms used in describing an interest profile is also inordinately large.

Since available thesauri do not utilize the inherent structure existing in a particular subject discipline, problems arise in both subject indexing of documents and subscriber interests. Although the effort involved in creating a comprehensive vocabulary is a large initial expenditure, significant savings in daily expenditures in indexing would accrue from such an effort.

2. Effectiveness of the SDI Notice as a Self-Sufficient Information Carrier

In this pilot test, the content of the SDI notice was obtained by duplicating selected portions of the article itself. Both the front and reverse sides of the notice could be utilized so that the information disseminated could occupy an area equivalent to that of two EAM-sized cards. Since information was disseminated only as it existed in the source document, an evaluation of the notice is really an evaluation of the quality of the extract.

The majority of source materials consisted of two types of journal articles, those appearing in trade journals, and those appearing in scholarly technical journals. In many of the technical journals, a concise abstract heads each article. In most trade journals the essence of an article existed only in an introductory paragraph or one buried in the text. The significant passages within each document were marked by the subject indexer for inclusion in the SDI notice. For technical reports, abstracts were extracted from the Technical Abstract Bulletin, published by Defense Documentation Center, and Scientific and Technical Aerospace Reports, published by National Aeronautics and Space Administration.

Almost all the information disseminated by this means was sufficient to indicate the subject content of the document or article. On the basis of reading the notice, participants could decide, in most cases, whether or not they would like to read the complete article.

In all cases, participants retained the SDI notice for future reference. Retention of the abstract indicates that the notice is a self-sufficient information carrier of some permanent value. Abstracts received most enthusiastically were of the informative type in which research results were reported in brief as well as an indication of the document content. Abstracts from technical journals were received least enthusiastically because of low information content. Some participants remarked that

trade journals included a high proportion of promotional material printed in the format of a technical article.

3. Effect of the SDI Notice on Influencing Information Acquisition Habits

In many instances, participants referred a notice to someone else. The direct transfer of information between the individuals might play a larger role in a future SDI system. In essence, the original recipient of a notice would act as a further switching point in channeling information to the right person. This phenomenon occurred most frequently for individuals working within the same branch or laboratory. However, some participants referred notices to individuals outside their own sphere of interest, which indicates an effective system of personal communication within Army Natick Laboratories.

Over the short period during the pilot test, participants still relied heavily on former sources of information. Also, participants were well aware of the temporary nature of the pilot test and thus did not plan to rely substantially on it. Since photocopies of journal articles and technical reports were not distributed automatically through SDI feedback, participants merely retained the notice for reference to the future use of the articles. The evaluation of the effect of SDI on personal information acquisition habits must await full-scale implementation in which provision will be made for automatic distribution of photocopied articles. An analysis could then be made of the relevancy of articles selected on the basis of the abstract. Other long-term studies that might be conducted are the effect on utilization of library information resources and the effect on interdisciplinary activities.

B. OPERATING COSTS

Operating costs are analyzed in terms of materials, equipment utilization, and personnel requirements. The total cost of operating the pilot SDI system is further analyzed into the unit cost per document processed, the unit cost per profile, and the unit cost for producing one SDI notice.

1. Raw Materials

The cost of raw materials is negligible compared with other unit costs of the system. The five materials used are 8 1/2 x 14" paper, 8 1/2 x 11" paper, Xerox card stock, EAM cards and rubber cement. The following table lists unit costs.

Paper (8 1/2 x 14")	$20.00 per carton (5,000 sheets)
Paper (8 1/2 x 11")	$16.85 per carton (5,000 sheets)
Xerox Card Stock	$75.00 per case (4,000 cards)
EAM Cards	$40.50 per carton (10,000 cards)
Rubber Cement	$ .45 per bottle

The Xerox card stock, purchased from Essex Systems, Inc. in New York City, is die cut sheets, each containing four EAM cards. The cards will pass through any low speed electronic accounting machine equipment.

2. Equipment Utilization

The highest single data processing cost was associated with computer usage. Approximately one minute is used in matching the interest profiles for twenty-five participants against the terms describing one document. Total time is proportional to the product of the total number of document terms and the total number of profile terms. For the fourth production cycle, 235 documents were processed in three hours.

The following is a table giving equipment utilization per work unit:

Keypunch	3.50 minutes per document
Card Sorter	.06 minutes per document
Card Sorter	.01 minutes per selection
Card Collator	.01 minutes per selection
Card Reproducer	.01 minutes per selection
Interpreter	.02 minutes per selection

A Xerox office copier was used to produce notice cards. In this operation, Xeroxing title pages and Tables of Contents required 1.54 minutes per document and Xeroxing notice cards required .32 minutes per selection.

3. Personnel Requirements

One trained subject indexer can index approximately thirty

documents per day if the subject terms are dictated into a recording machine. If the indexer is required to write out the terms on an indexing worksheet, the task can be accomplished at the rate of twenty documents per day. Including Xeroxing of notice cards, but excluding the computer matching process, approximately 1,000 notices could be processed in an eight-hour day by one person. During an eight-hour day, one clerk-typist could process approximately twenty-five documents, including the Xeroxing and stripping-up of title pages.

4. Cost Centers

For purposes of costing out the operation of the SDI system, three cost centers were developed. Document processing includes all those costs attributable to making the document available to the matching and selection function, and to notice duplication. Distribution processing includes all costs associated with matching, selecting, and processing selections. Profile costs are those associated with indexing the interests of subscribers and in preparing profiles in machine-readable form. Tables 2 and 3 give the costs for document and distribution processing. The third cost center, associated with profiling, can vary over a wide range and is very dependent upon the indexer's understanding of the subject matter and comprehension of the problems described by the participants. Typical unit costs cannot be extrapolated to other programs because of varying circumstances. During the course of the pilot test, approximately thirty man-days were used in profiling activities. Six man-days were consumed in personal and group interviews.

Compared with the time required to punch document subject cards and descriptive cataloging cards, keypunch requirements for profiles are minor. The number of cards required for each profile ranged from ten to over one hundred. Since the number of cards requiring re-keypunching due to profile changes is relatively small, the largest keypunching requirement is met during construction of the initial profile.

5. Proration of Costs Between Document Services

In a fully integrated system, the cost of document processing would be shared among other services. The same cards punched for the SDI system could also be used for generating accession lists, catalog cards, book-form indexes, and machine-searchable files. The costs incurred in subscriber profiling and document selection would be the only ones attributable directly to an SDI system.

TABLE 2
DOCUMENT PROCESSING

Task	Minutes Per Document
Log in journal	.01
Select documents and mark extracts	4.44
Type journal citation	.30
Xerox journal table of contents and document title page	1.54
Assign and stamp accession number and complete descriptive cataloging worksheet	1.54
Keypunch descriptive cataloging worksheet	3.00
Proof and correct title and journal citation	.27
Sort descriptive cataloging cards by accession number	.02
List descriptive cataloging cards	.01
Perform subject indexing	17.40
Keypunch subject terms	1.50
Proof and correct subject terms	.46
Sort subject terms by accession number	.08
Log out journal	.01
Cut document title page and strip-up	7.00
Total time	37.58 minutes

TABLE 3
DISTRIBUTION PROCESSING

Tasks	Minutes Per Selection
Match profile terms against document terms	.65
Sort copy at distribution cards	.01
List copy cards	.01
Xerox notice cards	.32
Burst notice cards	.20
Collate notice and copy cards	.04
Intersperse gang punch notice cards	.01
Sort notice cards	.01
Duplicate distribution cards into response cards	.01
Merge response and notice cards	.01
Intersperse gang punch notice cards	.01
Interpret response and notice cards	.02
Sort distribution and notice cards	.01
Total time	1.31 minutes

C. IMPLEMENTATION AT OTHER ARMY R&D FACILITIES

1. Adaptation of Computer Programs

Computer programs written for the pilot SDI system can be used with minor modifications at other installations. Three computer programs are available for maintaining a glossary of technical terms, recording document and profile terms on magnetic tapes, and matching the two resulting tapes to obtain SDI selections. Whether it is desirable to use these programs depends upon the current degree of mechanization of the adopting library. Highly-mechanized libraries will probably prefer to rewrite the programs so that the SDI system is closely integrated with other presently mechanized library services. On the other hand, libraries not currently mechanized will probably want to utilize these programs as a starting point. Since the objectives of the pilot test were broader than that of merely developing computer programs, users may also wish to modify the programs to increase their operating efficiency and thereby reduce the total computer utilization time.

2. System Modification to Meet Local Requirements

The method of producing the SDI notice card described in this report is recommended for three situations: a) ready availability of pre-abstracted materials, b) large number of copies (i.e., over ten) required for each document, and c) requirement to preserve information which cannot be keypunched. Any one or combination of these three circumstances would warrant the production of SDI notice cards by the method described in this report. In circumstances other than these, a conventional SDI system utilizing the on-line computer printer to produce notice images is more appropriate.

3. Development of a Technical Vocabulary

The degree of efficiency of the SDI matching function is directly proportional to the quality of the subject indexing tool and its application in the indexing of both subscriber profiles and documents. In the circumstances encountered at NLABS, a much higher efficiency could have been attained had an appropriate indexing tool been available. At some installations, it will be possible to utilize pre-existing externally produced thesauri. The adoption of this technique, however, should be tested on a small scale so that experience can be gained to gage the suitability of this approach.

4. Personnel Training

Very little training is required to implement and operate this SDI system. Assuming that librarians and data processing personnel will be responsible for carrying out the major processing operations, these personnel can learn their respective roles through reading the implementation manual which constitutes Volume II of this report. No specialized skills are required other than those normally encountered in document indexing and data processing. Clerk-typists can assume the bulk of work load involved in notice production. Clerical operations are restricted to typing, Xeroxing, cutting and stripping. If skilled technical typists are available, they can be utilized to assume some of the document indexing work load by transcribing technical terms dictated by the subject indexer.

In most installations, programs or plugboards are readily available to accomplish the interpreting, collating, and reproducing functions.

5. Interlaboratory Cooperation

As mentioned above, one of the conditions under which the proposed method of notice production is recommended is that of the availability of pre-existing abstracts. One of the most attractive features of this method is the universality of the graphic image. Data in machine-readable form usually must be converted before utilization in another system. Systems utilizing graphic reproduction of SDI notices, however, can share a common file of original printed information. Any image printed on an EAM card can be utilized with equal facility by all installations.

Although information contained in an abstract may be shared by different systems, the approach to subject indexing may differ significantly so that utilization of machine-readable index terms may not be possible. However, should the sharing of machine-readable data be intellectually feasible, the format for machine-readable processing in this system has been simplified so that the sharing of machine-readable data can be accomplished with minimal file conversion problems. Each document term is contained within a single record on the magnetic tape. This record contains the characters representing the term, the number of the subject indexing tool from which the term was extracted, and the document number.

D. INTERPRETATION OF PRODUCTION STATISTICS

Statistics on the four SDI production runs are presented in Appendixes A and B. Appendix A presents the statistics necessary to derive

the relationship between relevancy of selections and various parameters such as profile type, number of terms, and term source. Appendix B presents the distribution of the number of selections for each document.

1. Relevancy vs. Size of Profile

There does not seem to be a strong relationship between the number of terms in a profile and the degree of relevancy of selections. One relationship that does hold is that 75% of the profiles containing less than 20 terms are 50% relevant or better. The average percent of profiles achieving relevancy greater than 50% is respectively 64%, 30%, 48%, and 52% for the four production runs.

2. Relevancy vs. Glossary

Profiles indexed by the ITT Thesaurus yield the most relevant selections whereas those profiled by the EJC Thesaurus yield the least relevant documents. The validity of this conclusion is supported intuitively since it could be expected that, for NLABS interests, most relevant terms would be obtained from a textile glossary, whereas least relevant terms would be obtained from a general engineering glossary.

3. Relevancy vs. Type of Profile

A true relationship between the value of the weighted profile versus the Boolean profile cannot be determined from the test results because weighted profiles were used especially in difficult cases where the predicted number of matches was low. In the last production run, however, the average relevancy for weighted profiles was 39% and accounted for 50% of the 0% relevancy profiles (2 out of 4). Deleting weighted profile relevancies raises the average relevancy for the remaining profiles from 45% to 47%.

4. Man vs. Machine

On the last production cycle, an experiment was conducted to determine whether manual selection or computer selection yielded the highest relevancy ratios. The results of this experiment are summarized in the Table 4 entitled "Statistics of Manual Versus Computer Selection." The last two columns of this table interpret the meaning of the statistics presented. The column entitled "Which Method Missed More Relevant Documents," shows that a person selecting on the basis of document title alone misses more relevant documents than the computer selection process based on subject terms assigned by that same person. The second column entitled "Which

TABLE 4

STATISTICS OF MANUAL VS. COMPUTER SELECTION

Profile No.	Matches (Manual)	Of Interest (Manual)	% of Interest (Manual)	Matches (Machine)	Of Interest (Machine)	% of Interest (Machine)	Which Method Missed More Relevant Documents?	Which Method Added More Irrelevant Documents?
1	9	4	44	12	8	67	Man.	-
2	50	3	6	41	3	7	-	Man.
3	22	3	14	20	7	35	Man.	Man.
4	20	10	50	31	11	35	Man.	Comp.
5	17	15	88	24	18	75	Man.	Comp.
6	12	11	91	29	21	72	Man.	Comp.
7	9	1	11	4	3	75	Man.	Man.
8	2	1	50	2	1	50	-	-
9	1	1	100	2	0	0	Comp.	Comp.
10	8	4	50	1	0	0	Comp.	Man.
11	8	3	38	27	8	29	Man.	Comp.
12	26	26	100	16	16	100	Comp.	-
16	0	-	-	7	0	0	-	Comp.
17	28	4	14	36	6	17	Man.	Comp.
21	10	6	60	12	11	92	Man.	Man.
22	16	11	69	4	3	75	Comp.	Man.
23	9	7	78	12	7	58	-	Comp.
24	8	4	50	14	8	57	Man.	Comp.
25	0	-	-	0	-	-	-	-
26	4	3	75	7	4	57	Man.	Comp.
28	0	-	-	10	5	50	Man.	Comp.
29	46	15	33	68	19	28	Man.	Comp.
30	0	-	-	1	0	0	-	Comp.
31	0	-	-	0	-	-	-	-
32	10	8	80	13	6	46	Comp.	Comp.

Method Added More Irrelevant Documents," shows that the computer tends to select more irrelevant documents than the person selecting on the basis of title. This experiment indicates that the overall relevancy of an SDI system could be much enhanced by human pre-screening of the computer selected SDI notices. Average relevancy for manual selection was 55% as compared to 45% for computer selection. The difference in relevancy, however, is attributable to the larger number of irrelevant documents selected by the computer.

The fact that computer selection also misses relevant documents is supported by the last column of statistics on the first SDI cycle presented in Appendix A. This column labeled "Recall" shows the percentage of relevant documents missed, based on an estimation performed by each participant who scanned a total list of titles processed in the first SDI cycle.

5. Summary of Relevancy Statistics

Table 5 summarizes the relevancy statistics derived from Appendix A. This table shows that, for instance, the percentage of profiles achieving 50% relevancy or better, was 64% for Run I, 30% for Run II, 48% for Run III, and 52% for Run IV. The low relevancy for the second run was due to the addition of terms used to index documents judged relevant that were selected during Run I. This extreme dip in relevancy shows that terms from relevant documents cannot be used as a basis for profile readjustment.

Table 6 summarizes relevancy and number of relevant matches for each subscriber. These figures are an extraction from Appendix A.

6. Document Selection Pattern

Appendix B shows the distribution of the number of times documents were selected in each of the production cycles. These statistics can be used to plan reproduction processing since they show how many documents were selected once, twice, three times, etc. A significant factor is the number of documents selected 0 tiems, since this indicates the relative match of the document collection to the needs of the subscribers. For the four production cycles, the numbers of documents selected 0 times were respectively 15 out of 111 processed, 186 out of 418 processed, 86 out of 247 processed, and 88 out of 235 processed. Those documents selected two or three times accounted for the highest percentage of notices.

TABLE 5

PERCENT RELEVANCY VS. NUMBER OF RELEVANT MATCHES

Profile No.	Run I	Run II	Run III	Run IV
1	50/7	29/2	40/4	67/8
2	36/9	8/5	22/8	7/3
3	50/2	27/20	21/6	35/7
4	53/10	21/16	30/12	35/11
5	47/7	41/17	83/20	75/18
6	83/15	56/44	50/14	72/21
7	50/1	0/0	100/1	75/3
8	-/-	41/25	0/0	50/1
9	25/9	0/0	0/0	0/0
10	71/15	42/13	60/3	0/0
11	10/2	4/1	21/7	29/8
12	20/7	87/46	100/7	100/16
16	50/1	40/2	0/0	0/0
17	42/8	12/8	10/4	17/6
21	71/17	72/39	73/11	92/11
22	30/3	90/9	0/0	75/3
23	69/9	56/23	50/2	58/7
24	70/14	-/-	57/12	57/8
25	-/-	55/6	-/-	-/-
26	64/9	31/13	51/5	57/4
28	100/3	21/6	60/3	50/5
29	66/25	37/23	33/22	28/19
30	33/1	60/3	100/1	20/0
31	-/-	-/-	-/-	-/-
32	90/4	33/2	18/2	46/6

TABLE 6

SUMMARY OF RELEVANCY STATISTICS

		Run I	Run II	Run III	Run IV
Average relevancy		54	37	43	45
Average of profiles achieving over 10% relevancy		56	45	54	57
Percent profiles achieving relevancy greater than or equal to:	90%	9	4	13	9
	80%	14	9	17	9
	70%	27	13	22	26
	60%	41	17	30	30
	50%	64	30	48	52
	40%	73	48	52	57
	30%	86	61	61	65
	20%	95	78	74	74
	10%	100	83	83	78
	0%	100	100	100	100

VI. RECOMMENDATIONS AND CONCLUSIONS

A. RECOMMENDATIONS

1. Before implementation of a full-scale SDI system, each installation should carefully analyze its subject indexing requirements. If a controlled vocabulary does not already exist, existing vocabularies can be profitably tested using the model described in this pilot test.

2. Three criteria should be used in determining SDI reproduction requirements. These criteria are: (a) need to reproduce formulae and other graphic representations, (b) medium to high notice copy requirements, (c) availability of pre-existing abstracts.

3. An SDI system should be integrated with other mechanized information services so that multiple use can be made of document data currently put in machine-readable form.

4. An SDI system should be implemented in stages to acquaint the users and the operators with the methods and procedures employed and to uncover special problems that may exist.

5. Full-scale implementation should utilize inverted files to reduce computer matching time.

6. Indexers should be employed who thoroughly understand the intricacies of the subject being indexed.

7. Maximum use should be made of document data already in machine-readable form from external sources.

8. This model system should continue to be used as a research tool in discovering additional relationships that may exist between SDI parameters.

B. CONCLUSIONS

1. A potential exists for deriving highly relevant profiles constructed by a combination of the Boolean and Weighted match strategies.

2. When subject interests are diverse and, at the same time, in-depth, a hierarchical structuring of vocabulary terms is necessary to achieve high relevancy at an efficient level.

3. Matching on inverted files yields higher processing efficiency.

4. Close contact with the system's users is extremely necessary, both from the standpoint of communications and personal relations.

5. There is an inverse relationship between the number of selections made by the computer and the percentage of relevant selections.

6. A strong relationship does not exist between the number of terms in the profile and the relevancy of selections.

7. Pre-existing abstracts for technical journals and Government reports are sufficiently informative to be used on the SDI notice, without modification.

APPENDIX A

SDI PRODUCTION STATISTICS
RUNS I THROUGH IV

	Profile No.	Profile Type	Levels/Profile	Glossary	No. of Terms	Of Interest Want Document	Of Interest Don't Want Document	Of Interest Have Seen	Total of Interest	Of No Interest	No Response	Notices	% of Interest	Titles Missed	Total of Interest	Recall	
1	1	B	1	ITT	11	3	3	1	7	7	0	14					
	2	B	1	ITT	11	-	-	-	-	-	-	0					
TOTAL					22	3	3	1	7	7	0	14	50	0	7	100	
2	1	B	1	ITT	29	1	1	0	2	6	0	8					
	2	B	1	EJC	10	2	5	0	7	10	0	17					
TOTAL					39	3	6	0	9	16	0	25	36	6	15	50	
3	1	B	1	ITT	30	2	0	0	2	2	0	4	50	7	9	22	
4	1	B	1	ITT	42	2	8	0	10	9	0	19					
	2	B	1	ITT	4	-	-	-	-	-	-	0					
TOTAL					46	2	8	0	10	9	0	19	53	3	13	50	
5	1	B	1	ITT	26	1	6	0	7	8	0	15	47	1	8	88	
6	1	B	1	ITT	14	4	2	3	9	0	0	9					
	2	B	1	ITT	10	2	3	1	6	3	0	9					
TOTAL					24	6	5	4	15	3	0	18	83	7	22	68	
7	1	B	1	DDC	17	1	0	0	1	1	0	2	50	0	1	100	
8	1	B	1	ITT	14	-	-	-	-	-	-	0	-	7	7	0	
9	1	B	1	EJC		8	0	1	9	3	0	4	25	0	1	100	
10	1	B	1	EJC	8	4	9	0	13	5	0	18					
	2	B	1	DDC	13	0	2	0	2	1	0	3					
TOTAL					21	4	11	0	15	6	0	21	71	2	17	58	
11	1	B	1	DDC	30	0	1	1	2	19	0	21	10	2	4	50	
12	1	B	1	DDC	23	2	1	0	3	4	0	7					
	2	B	1	DDC	27	3	0	0	3	22	0	25					
	3	B	1	DDC	5	-	-	-	-	-	-	-					
	4	B	1	DDC	4	0	0	0	0	2	0	2					
	5	B	1	DDC	5	1	0	0	1	0	0	1					
TOTAL					64	6	1	0	7	28	0	35	20	7	14	50	
16	1	B	1	DDC	12	0	0	1	1	1	0	2	50	1	2	50	
17	1	B	1	EJC	48	1	7	0	8	11	0	19	42	6	14	57	
21	1	B	1	ITT	6	1	1	2	4	0	0	4					
	2	B	1	EJC	2	0	0	0	0	0	0	0					
	3	B	1	DDC	5	1	12	0	13	7	0	20					
TOTAL					13	2	13	2	17	7	0	24	71	4	21	81	
22	1	B	1	DDC	11	0	0	0	0	1	0	1					
	2	B	1	ITT	13	0	1	1	2	3	0	5					
	3	B	1	DDC	9	-	-	-	-	-	0	0					
	4	B	1	DDC	11	-	-	-	-	-	0	0					
	5	B	1	DDC	3	1	0	0	1	3	0	4					
	6	B	1	DDC	6	-	-	-	-	-	0	0					
TOTAL					53	1	1	1	3	7	0	0	30	2	5	60	

	Profile No.	Profile Type	Levels/Profile	Glossary	No. of Terms	Of Interest Want	Don't Want	Have Seen	Total of Interest	No Interest	No Response	Notices	% of Interest	Titles Missed	Total of Interest Recall	
23	1	B	1	EJC	14	0	7	2	9	4	0	13	69	1	10	90
24	1	B	1	EJC	12	0	9	0	9	5	0	14				
	2	B	1	EJC	7	-	-	-	-	-	0	0				
	3	B	1	EJC	8	-	-	-	-	-	0	0				
	4	B	1	EJC	9	-	-	-	-	-	0	0				
	5	B	1	DDC	9	4	1	0	5	1	0	6				
TOTAL					45	4	10	0	14	6	0	20	70	5	15	93
25	1	B	1	EJC	49	-	-	-	-	-	-	0	-	9	9	0
26	1	B	1	EJC	21	4	5	0	9	5	0	4	64	2	11	82
28	1	B	1	EJC	22	3	0	0	3	0	0	3	100	4	7	43
29	1	B	1	EJC	15	3	1	2	6	7	0	13				
	2	B	1	ITT	5	1	0	0	1	0	0	1				
	3	B	1	EJC	4	1	0	0	1	0	0	1				
	4	B	1	EJC	7	2	2	0	4	2	0	6				
	5	B	1	ITT	15	10	1	0	11	4	0	15				
	6	B	1	ITT	15	2	0	0	2	0	0	2				
TOTAL					61	19	4	2	25	13	0	38	66	3	28	89
30	1	B	1	EJC	43	0	1	0	1	2	0	3	33	3	4	25
31	1	B	1	DDC	29	-	-	-	-	-	2	2	-	-	-	-
32	1	B	1	EJC	8	-	-	-	-	-	0	0				
	2	B	1	EJC	11	0	2	0	2	1	0	3				
	3	B	1	ITT	11	2	0	0	2	0	0	2				
	4	B	1	EJC	3	-	-	-	-	-	0	0				
TOTAL					33	2	2	0	4	1	0	5	90	2	6	67

Subscriber #	Profile No.	Profile Type	Levels/Profile	ITT Glossary	EJC Glossary	DDC Glossary	No. of Terms	Of Interest Want Document	Of Interest Don't Want Document	Of Interest Have Seen	Total of Interest	Of No Interest	No Response	Notices	% of Interest
1	1	C	3	21	25	0	46	1	0	1	2	5	0	7	
	2	C	2	13	0	0	13	-	-	-	-	-	-	0	
TOTAL				33	25	0	59	1	0	1	2	5	0	7	29
2	1	C	2	18	9	7	34	0	0	0	0	1	0	1	
	2	C	2	28	31	15	74	0	5	0	5	53	1	59	
TOTAL				46	40	22	108	0	5	0	5	54	1	60	8
3	1	C	2	59	12	11	82	4	16	0	20	53	0	73	27
4	1	C	2	97	0	11	108	9	7	0	16	61	0	77	
	2	C	2	4	0	0	4	-	-	-	-	-	-	0	
TOTAL				101	0	11	112	9	7	0	16	61	0	77	21
5	1	C	2	82	0	0	82	0	17	0	17	24	0	41	41
6	1	C	2	0	63	31	94	3	34	2	39	27	2	68	
	2	C	2	39	0	0	39	0	4	1	5	6	0	11	
TOTAL				39	63	31	133	3	38	3	44	33	2	79	56
7	1	C	2	0	0	29	29	0	0	0	0	9	0	9	
	2	C	2	0	8	0	8	0	0	0	0	19	0	19	
TOTAL				0	8	29	37	0	0	0	0	28	0	28	0
8	1	C	2	14	33	23	70	4	19	2	25	33	3	61	
	2	C	2	0	5	0	5	-	-	-	-	-	-	0	
TOTAL				14	38	23	75	4	19	2	25	33	3	61	41
9	1	C	2	0	16	5	21	0	0	0	0	6	0	6	0
10	1	C	3	0	34	10	44	3	9	1	13	17	1	31	
	2	C	2	0	1	10	11	-	-	-	-	-	-	0	
TOTAL				0	35	20	55	3	9	1	13	17	1	31	42
11	1	W	1	0	16	30	46	1	0	0	1	20	4	25	4
12	1	C	2	0	0	39	39	2	5	0	7	3	0	10	
	2	C	2	0	30	23	53	13	26	0	39	4	0	43	
	3	C	2	0	0	5	5	-	-	-	-	-	-	0	
	4	C	2	0	0	4	4	-	-	-	-	-	-	0	
	5	C	2	0	0	5	5	-	-	-	-	-	-	0	
	6	C	2	11	0	0	11	-	-	-	-	-	-	0	
	7	C	2	0	0	3	3	-	-	-	-	-	-	0	
TOTAL				11	30	79	120	25	31	0	46	7	0	53	87
16	1	W	1	0	5	19	24	1	1	0	2	3	0	5	40
17	1	C	2	0	78	16	94	5	3	0	8	57	0	65	12

Subscriber #	Profile No.	Profile Type	Levels/Profile	ITT Glossary	EJC Glossary	DDC Glossary	No. of Terms	Of Interest Want Document	Of Interest Don't Want Document	Of Interest Have Seen	Total of Interest	Of No Interest	No Response	Notices	% of Interest
21	1	C	2	6	0	0	6	0	1	0	1	0	0	1	
	2	C	2	0	2	0	2	-	-	-	-	-	-	0	
	3	C	2	0	0	5	5	1	1	0	2	0	0	2	
	4	C	2	22	0	0	22	0	0	2	2	2	0	4	
	5	C	2	0	14	0	14	6	16	0	22	1	0	23	
	6	C	2	0	0	28	28	0	10	0	10	4	0	14	
	7	C	2	0	0	20	20	0	2	0	2	8	0	10	
TOTAL				28	16	53	97	7	30	2	39	15	0	54	77
22	1	C	2	0	0	11	11	2	0	0	2	0	0	2	
	2	C	2	24	0	0	24	0	1	1	2	0	0	2	
	3	C	2	0	0	9	9	-	-	-	-	-	-	0	
	4	C	2	0	0	11	11	-	-	-	-	-	-	0	
	6	C	2	0	0	6	6	-	-	-	-	-	-	0	
	7	C	2	0	6	0	6	1	4	0	5	1	0	6	
	8	C	2	0	0	4	4	-	-	-	-	-	-	0	
	9	C	2	0	5	0	5	-	-	-	-	-	-	0	
TOTAL				24	11	41	76	3	5	1	9	1	0	10	90
23	1	C	2	0	37	6	43	0	19	4	23	18	0	41	56
24	1	C	3	0	12	5	17	0	0	0	0	0	4	4	
	2	C	2	0	7	0	7	-	-	-	-	-	-	0	
	3	C	2	0	8	0	8	-	-	-	-	-	-	0	
	4	C	2	0	9	0	9	-	-	-	-	-	-	0	
	5	C	2	0	0	12	12	-	-	-	-	-	-	0	
	6	C	2	0	10	19	29	0	0	0	0	0	5	5	
	7	C	3	11	13	8	32	0	0	0	0	0	1	1	
TOTAL				11	59	44	114	0	0	0	0	0	10	10	-
25	1	C	2	0	49	0	49	-	-	-	-	-	-	0	
	2	C	2	0	0	26	26	1	1	2	4	0	0	4	
	3	C	2	0	7	0	7	1	1	0	2	5	0	7	
TOTAL				0	56	26	82	2	2	2	6	5	0	11	55
26	1	C	2	0	54	1	55	4	8	0	12	27	2	41	
	2	C	2	11	0	0	11	-	-	-	-	-	-	0	
	3	C	2	0	0	4	4	1	0	0	1	0	0	1	
TOTAL				11	54	5	70	5	8	0	13	27	2	42	31
28	1	W	1	0	22	0	22	-	-	-	-	-	-	0	
	2	C	2	0	22	0	22	4	0	0	4	22	0	26	
	3	C	2	0	0	8	8	-	-	-	-	-	-	0	
	4	C	2	0	0	6	6	2	0	0	2	0	0	2	
TOTAL				0	44	14	58	6	0	0	6	22	0	28	21

STATISTICS ON FIRST SDI CYCLE (CONT'D)

Subscriber #	Profile No.	Profile Type	Levels/Profile	ITT Glossary	EJC Glossary	DDC Glossary	No. of Terms	Of Interest Want Document	Of Interest Want Document	Of Interest Don't Have Seen	Total of Interest	Of No Interest	No Response	Notices	% of Interest
29	1	C	2	0	15	8	23	5	0	1	6	13	0	19	
	2	C	2	8	0	0	8	-	-	-	-	-	-	0	
	3	C	2	0	4	0	4	-	-	-	-	-	-	0	
	4	C	2	0	50	0	50	8	1	1	10	26	0	36	
	5	C	2	27	0	0	27	3	1	1	5	1	0	6	
	6	C	2	15	0	0	15	-	-	-	-	-	-	0	
	7	C	2	13	0	0	13	1	0	0	1	0	0	1	
	8	C	2	8	0	0	8	-	-	-	-	-	-	0	
	9	C	2	6	0	0	6	0	1	0	1	0	0	1	
	10	C	2	8	0	0	8	-	-	-	-	-	-	0	
TOTAL				85	69	8	162	17	3	3	23	40	0	63	37
30	1	W	1	0	48	19	67	0	3	0	3	2	0	5	60
31	1	W	1	0	0	29	29	-	-	-	-	-	-	0	-
32	1	C	2	0	8	0	8	-	-	-	-	-	-	0	
	2	C	2	0	11	0	11	-	-	-	-	-	-	0	
	3	C	2	16	3	0	19	0	0	1	1	3	0	4	
	5	C	2	5	7	0	12	2	0	0	2	0	0	2	
	6	C	2	11	0	0	11	0	0	0	0	2	0	2	
	7	C	2	2	0	0	2	0	0	0	0	1	0	1	
	8	C	2	0	12	0	12	-	-	-	-	-	-	0	
TOTAL				34	41	0	75	2	0	1	3	6	0	9	33

Subscriber #	Profile No.	Profile Type	Levels/Profile	ITT Glossary	EJC Glossary	DDC Glossary	No. of Terms	Of Interest Want Document	Of Interest Don't Want Document	Of Interest Have Seen	Total of Interest	Of No Interest	No Response	Notices	% of Int.
1	1	C	2	47	0	0	47	3	1	0	4	6	0	10	40
2	1	C	2	27	27	12	66	0	8	0	8	28	1	37	22
3	1	C	2	45	13	10	68	0	6	0	6	23	0	29	21
4	1	C	2	82	0	2	84	1	11	0	12	28	0	40	30
5	1	C	2	71	0	0	71	1	19	0	20	4	0	24	83
6	1	C	2	17	28	11	56	0	14	0	14	14	0	28	50
7	1	C	3	0	0	18	18	0	0	1	1	0	0	1	100
8	1	C	2	12	21	15	48	0	0	0	0	1	0	1	0
9	1	C	2	0	8	8	16	0	0	0	0	2	0	2	0
10	1	W	1	3	3	3	9	1	2	0	3	2	0	5	60
11	1	W	1	0	11	19	30	7	0	0	7	26	0	33	21
12	1	C	2	5	10	12	27	3	1	0	4	0	0	4	
	2	C	2	0	16	19	35	2	1	0	3	0	0	3	
	3	W	1	0	0	5	5	0	0	0	0	0	0	0	
	4	C	2	0	0	4	4	0	0	0	0	0	0	0	
	5	C	2	0	0	5	5	0	0	0	0	0	0	0	
TOTAL				5	26	45	76	5	2	0	7	0	0	7	100
16	1	C	2	0	17	21	38	0	0	0	0	6	0	6	0
17	1	C	2	0	27	2	29	4	0	0	4	37	0	41	10
21	1	W	1	15	0	0	15	1	10	0	11	4	0	15	73
22	1	C	2	17	7	38	62	0	0	0	0	1	0	1	0
23	1	W	1	0	12	6	18	2	0	0	2	2	0	4	50
24	1	W	1	4	20	12	36	6	2	0	8	2	0	10	
	2	C	2	5	17	17	39	0	4	0	4	7	0	11	
TOTAL				9	37	29	75	6	6	0	12	9	0	21	57
25	1	C	2	0	48	20	68	0	0	0	0	0	0	0	-
26	1	C	2	3	20	4	27	3	2	0	5	3	1	9	51
28	1	C	2	0	30	11	41	3	0	0	3	2	0	5	60
29	1	C	2	78	57	4	139	15	0	7	22	45	0	67	33
30	1	W	1	0	43	19	62	1	0	0	1	0	0	1	100
31	1	W	1	0	19	38	57	0	0	0	0	0	0	0	-
32	1	W	1	13	27	0	40	0	2	0	2	9	0	11	18

TOTAL 1260

Subscriber #	Profile No.	Profile Type	Levels/Profile	ITT Glossary	EJC Glossary	DDC Glossary	No. of Terms	Of Interest Want Document	Of Interest Don't Want Document	Of Interest Have Seen	Total of Interest	Of No Interest	No Response	Notices	% Of Int..
1	1	C	2	47	0	0	47	2	6	0	8	4	0	12	67
2	1	C	2	27	27	12	66	0	3	0	3	38	0	41	7
3	1	C	2	45	13	10	68	1	6	0	7	13	0	20	35
4	1	C	2	82	0	2	84	1	10	0	11	20	0	31	35
5	1	C	2	71	0	0	71	2	16	0	18	6	0	24	75
6	1	C	2	17	28	11	56	0	17	4	21	8	0	29	72
7	1	C	3	0	0	18	18	0	0	3	3	1	0	4	75
8	1	C	2	12	21	15	48	0	1	0	1	1	0	2	50
9	1	C	2	0	8	8	16	0	0	0	0	2	0	2	0
10	1	W	1	3	3	3	9	0	0	0	0	1	0	1	0
11	1	W	1	0	11	19	30	5	0	3	8	18	1	27	29
12	1	C	2	5	10	12	27	3	1	0	4	0	0	4	
	2	C	2	0	16	19	35	7	5	0	12	0	0	12	
	3	W	1	0	0	5	5	-	-	-	-	-	-	0	
	4	C	2	0	0	4	4	-	-	-	-	-	-	0	
	5	C	2	0	0	5	5	-	-	-	-	-	-	0	
TOTAL				5	26	45	76	10	6	0	16	0	0	16	100
16	1	C	2	0	17	21	38	0	0	0	0	7	0	7	0
17	1	C	2	0	27	2	29	5	0	1	6	30	0	36	17
21	1	W	1	15	0	0	15	2	5	4	11	1	0	12	92
22	1	C	2	17	7	38	62	2	0	1	3	1	0	4	75
23	1	W	1	0	12	6	18	0	1	6	7	5	0	12	58
24	1	W	1	4	20	12	36	-	-	-	-	-	-	0	
	2	C	2	5	17	17	39	4	4	0	8	6	0	14	
TOTAL				9	37	29	75	4	4	0	8	6	0	14	57
25	1	C	2	0	48	20	68	-	-	-	-	-	-	0	-
26	1	C	2	3	20	4	27	3	1	0	4	3	0	7	57
28	1	C	2	0	30	11	41	5	0	0	5	5	0	10	50
29	1	C	2	78	57	4	139	19	0	0	19	49	0	68	28
30	1	W	1	0	43	19	62	0	0	0	0	1	0	1	0
31	1	W	1	0	19	38	57	-	-	-	-	-	-	0	-
32	1	W	1	13	27	0	40	6	0	0	6	7	0	13	46
TOTAL							1260							58	

STATISTICS ON FOURTH SDI CYCLE

APPENDIX B

DOCUMENT SELECTION PATTERNS

PRODUCTIVE CYCLES I THROUGH IV

DOCUMENT SELECTION PATTERN FOR FIRST PRODUCTION CYCLE

15	Documents	were	selected	0	times,	giving	0	notices
19	"	"	"	1	"	"	19	"
13	"	"	"	2	"	"	26	"
29	"	"	"	3	"	"	87	"
8	"	"	"	4	"	"	32	"
12	"	"	"	5	"	"	60	"
4	"	"	"	6	"	"	24	"
8	"	"	"	7	"	"	56	"
3	"	"	"	8	"	"	24	"
111	Documents	were	processed	giving			328	notices

Notices	% Total	Cumulative %	Documents	% Total	Cumulative %
87	26.5	26.5	29	26.1	26.1
60	18.3	44.8	12	10.8	36.9
56	17.1	61.9	8	7.2	44.1
32	9.8	71.7	8	7.2	51.3
26	7.9	79.6	13	11.7	63.0
24	7.3	86.9	4	3.6	66.6
24	7.3	94.2	3	2.7	69.3
19	5.8	100.0	19	17.1	85.4
0	0	100.0	15	13.5	99.9
328	100.0		111	99.9	

DOCUMENT SELECTION PATTERN FOR SECOND PRODUCTION CYCLE

186	Documents	were	selected	0	times,	giving	0	notices
71	"	"	"	1	"	"	71	"
44	"	"	"	2	"	"	88	"
36	"	"	"	3	"	"	108	"
19	"	"	"	4	"	"	76	"
10	"	"	"	5	"	"	50	"
9	"	"	"	6	"	"	54	"
8	"	"	"	7	"	"	56	"
7	"	"	"	8	"	"	56	"
7	"	"	"	9	"	"	63	"
1	"	"	"	10	"	"	10	"
6	"	"	"	11	"	"	66	"
3	"	"	"	12	"	"	36	"
4	"	"	"	13	"	"	52	"
6	"	"	"	14	"	"	84	"
0	"	"	"	15	"	"	0	"
1	"	"	"	16	"	"	16	"
418	Documents	were	processed	giving			886	notices

Notices	% Total	Cumulative %	Documents	% Total	Cumulative %
108	12.2	12.2	36	8.6	8.6
88	9.9	22.1	44	10.5	19.1
84	9.5	31.6	6	1.4	20.5
76	8.6	40.2	19	4.5	25.0
71	8.0	48.2	71	17.0	42.0
66	7.5	55.7	6	1.4	43.4
63	7.1	62.8	7	1.7	45.1
56	6.3	69.1	8	1.9	47.0
56	6.3	75.4	7	1.7	48.7
54	6.1	81.5	9	2.2	50.9
52	5.9	87.4	4	1.0	51.9
50	5.6	93.0	10	2.4	54.3
36	4.1	97.1	3	0.7	55.0
16	1.8	98.9	1	0.2	55.2
10	1.1	100.0	1	0.2	55.4
0	0	100.0	186	44.5	99.9
886	100.0		418	99.9	

DOCUMENT SELECTION PATTERN FOR THIRD PRODUCTION CYCLE

86	Documents	were	selected	0	times,	giving	0 notices
50	"	"	"	1	"	"	50 "
52	"	"	"	2	"	"	104 "
20	"	"	"	3	"	"	60 "
19	"	"	"	4	"	"	76 "
14	"	"	"	5	"	"	70 "
4	"	"	"	6	"	"	24 "
2	"	"	"	7	"	"	14 "
247	Documents	were	processed	giving			398 notices

Notices	% Total	Cumulative %	Documents	% Total	Cumulative %
104	26.1	26.1	52	21.0	21.0
76	19.1	45.2	19	7.7	28.7
70	17.6	62.8	14	5.7	34.4
60	15.1	77.9	20	8.1	42.5
50	12.6	90.5	50	20.2	62.7
24	6.0	96.5	4	1.6	64.3
14	3.5	100.0	2	.8	65.1
0	0	100.0	86	34.8	99.9
398	100.0		247	99.9	

DOCUMENT SELECTION PATTERN FOR FOURTH PRODUCTION CYCLE

88	Documents	were	selected	0	times,	giving	0	notices
47	"	"	"	1	"	"	47	"
37	"	"	"	2	"	"	74	"
19	"	"	"	3	"	"	57	"
18	"	"	"	4	"	"	72	"
17	"	"	"	5	"	"	85	"
7	"	"	"	6	"	"	42	"
1	"	"	"	7	"	"	7	"
0	"	"	"	8	"	"	0	"
1	"	"	"	9	"	"	9	"
235	Documents	were	processed	giving			393	notices

Notices	% Total	Cumulative %	Documents	% Total	Cumulative %
85	21.6	21.6	17	7.2	7.2
74	18.8	40.4	37	15.8	23.0
72	18.3	58.7	18	7.6	30.6
57	14.5	73.2	19	8.1	38.7
47	12.0	85.2	47	20.0	58.7
42	10.7	95.9	7	3.0	61.7
9	2.3	98.2	1	.4	62.1
7	1.8	100.0	1	.4	62.5
0	0	100.0	88	37.5	100.0
393	100.0		235	100.0	

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) INFORMATION DYNAMICS CORPORATION 80 Main Street Reading, Massachusetts 01867	2a REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b GROUP NA

3. REPORT TITLE

Selective Dissemination of Information: Volume I - Pilot Test At U. S. Army Natick Laboratories

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

Final Report

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

Bivona, William A.

6. REPORT DATE May 1967	7a. TOTAL NO OF PAGES	7b NO OF REFS None
8a CONTRACT OR GRANT NO. DA 19-129-AMC0957 (N) b. PROJECT NO. c d	9a. ORIGINATOR'S REPORT NUMBER(S) IDC 8074 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) ATLIS Report No. 15	

10 AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited

11 SUPPLEMENTARY NOTES This report was prepared for the Army Technical Library Improvement Studies (ATLIS) Program	12. SPONSORING MILITARY ACTIVITY U. S. Army Natick Laboratories Natick, Massachusetts

13. ABSTRACT This report presents the results of a nine-month test of a prototype SDI system developed for Army Technical Libraries. During the pilot test, one thousand documents were cataloged and indexed and disseminated to twenty-five scientific and technical personnel at U. S. Army Natick Laboratories. During the course of the pilot test, statistics were accumulated on operating costs and various parameters affecting the relevancy of disseminated information. The prototype SDI system utilizes graphic techniques for producing multiple copies of each document abstract. The output of the computerized matching function is a set of EAM cards which are subsequently used to collate reproduced abstracts for dissemination. Collation is accomplished by interspersed gangpunching, merging, and sorting techniques.

DD FORM 1 JAN 64 1473

Security Classification

Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Documentation						
Distribution						
Bibliographies						
Programming (Computers)						
Dissemination						
Information Retrieval						
Reviews						
Design						
Economics						
History						
Selection Rules						
SDI (Identifier)						
Analysis						
ATLIS Program (Identified)						

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.

2a. **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.

2b. **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.

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

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

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

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

9a. **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.

9b. **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.

Security Classification