NITED STATES
ARTMENT OF
MMERCE
BLICATION

NBS TECHNICAL NOTE 599

A Survey of Selected Document Processing Systems

U.S.
PARTMENT
OF
OMMERCE
National
Bureau
of
ndards

NATIONAL BUREAU OF STANDARDS

The National Bureau of Standards[1] was established by an act of Congress March 3, 1901. The Bureau's overall goal is to strengthen and advance the Nation's science and technology and facilitate their effective application for public benefit. To this end, the Bureau conducts research and provides: (1) a basis for the Nation's physical measurement system, (2) scientific and technological services for industry and government, (3) a technical basis for equity in trade, and (4) technical services to promote public safety. The Bureau consists of the Institute for Basic Standards, the Institute for Materials Research, the Institute for Applied Technology, the Center for Computer Sciences and Technology, and the Office for Information Programs.

THE INSTITUTE FOR BASIC STANDARDS provides the central basis within the United States of a complete and consistent system of physical measurement; coordinates that system with measurement systems of other nations; and furnishes essential services leading to accurate and uniform physical measurements throughout the Nation's scientific community, industry, and commerce. The Institute consists of a Center for Radiation Research, an Office of Measurement Services and the following divisions:

Applied Mathematics—Electricity—Heat—Mechanics—Optical Physics—Linac Radiation[2]—Nuclear Radiation[2]—Applied Radiation[2]—Quantum Electronics[3]—Electromagnetics[3]—Time and Frequency[3]—Laboratory Astrophysics[3]—Cryogenics[3].

THE INSTITUTE FOR MATERIALS RESEARCH conducts materials research leading to improved methods of measurement, standards, and data on the properties of well-characterized materials needed by industry, commerce, educational institutions, and Government; provides advisory and research services to other Government agencies; and develops, produces, and distributes standard reference materials. The Institute consists of the Office of Standard Reference Materials and the following divisions:

Analytical Chemistry—Polymers—Metallurgy—Inorganic Materials—Reactor Radiation—Physical Chemistry.

THE INSTITUTE FOR APPLIED TECHNOLOGY provides technical services to promote the use of available technology and to facilitate technological innovation in industry and Government; cooperates with public and private organizations leading to the development of technological standards (including mandatory safety standards), codes and methods of test; and provides technical advice and services to Government agencies upon request. The Institute also monitors NBS engineering standards activities and provides liaison between NBS and national and international engineering standards bodies. The Institute consists of the following technical divisions and offices:

Engineering Standards Services—Weights and Measures—Flammable Fabrics—Invention and Innovation—Vehicle Systems Research—Product Evaluation Technology—Building Research—Electronic Technology—Technical Analysis—Measurement Engineering.

THE CENTER FOR COMPUTER SCIENCES AND TECHNOLOGY conducts research and provides technical services designed to aid Government agencies in improving cost effectiveness in the conduct of their programs through the selection, acquisition, and effective utilization of automatic data processing equipment; and serves as the principal focus within the executive branch for the development of Federal standards for automatic data processing equipment, techniques, and computer languages. The Center consists of the following offices and divisions:

Information Processing Standards—Computer Information—Computer Services—Systems Development—Information Processing Technology.

THE OFFICE FOR INFORMATION PROGRAMS promotes optimum dissemination and accessibility of scientific information generated within NBS and other agencies of the Federal Government; promotes the development of the National Standard Reference Data System and a system of information analysis centers dealing with the broader aspects of the National Measurement System; provides appropriate services to ensure that the NBS staff has optimum accessibility to the scientific information of the world, and directs the public information activities of the Bureau. The Office consists of the following organizational units:

Office of Standard Reference Data—Office of Technical Information and Publications—Library—Office of Public Information—Office of International Relations.

---

[1] Headquarters and Laboratories at Gaithersburg, Maryland, unless otherwise noted; mailing address Washington, D.C. 20234.
[2] Part of the Center for Radiation Research.
[3] Located at Boulder, Colorado 80302.

UNITED STATES DEPARTMENT OF COMMERCE
Maurice H. Stans, Secretary
NATIONAL BUREAU OF STANDARDS • Lewis M. Branscomb, Director

NBS TECHNICAL NOTE 599

ISSUED OCTOBER 1971
Nat. Bur. Stand. (U.S.), Tech. Note 599, 67 pages (October 1971)
CODEN: NBTNA

A Survey of Selected Document Processing Systems

Elizabeth Fong

Systems Development Division
Center for Computer Sciences and Technology
National Bureau of Standards
Washington, D.C. 20234

NBS Technical Notes are designed to supplement the Bureau's regular publications program. They provide a means for making available scientific data that are of transient or limited interest. Technical Notes may be listed or referred to in the open literature.

For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 20402.
(Order by SD Catalog No. C 13.46:599). Price 65 cents.

TABLE OF CONTENTS

A SURVEY OF SELECTED DOCUMENT PROCESSING SYSTEMS*

Elizabeth Fong

There are many document processing systems that are commercially available or government-owned. These systems emerged in the evolution from early efforts in library automation to current on-line systems. Due to the diverse nature of the facilities provided in the document processing systems, it is difficult to evaluate them. The purpose of this paper is to present a list of features as a set of dimensions along which to compare the surveyed systems. The feature list is also developed to serve as a common basis for describing document processing systems. Another purpose of this paper is to provide a reference tool for the eight systems surveyed. They are CIRCOL, DDC, ITIRC, The Mead Data Central, MEDLARS II, New York Times Information Bank, ORBIT II, and RECON/STIM. This paper first explores the characteristics of available, large document processing systems in general. An overview of the eight systems surveyed is presented. The paper then defines the feature list. The description of the eight systems surveyed according to the feature list outline is included as an Appendix.

Key words: Bibliographic system; computer package; data base; document processing; information retrieval; document storage and retrieval; text processing.

I. INTRODUCTION

Document processing syztems, sometimes referred to as document storage and retrieval systems, are computer-based systems that perform the function of a library, technical information center, or filing cabinet. Berul [1][1] defines a document processing system as a system that searches a collection of documents and delivers the documents or references most likely to be relevant. Question-answering or fact retrieval systems generate a direct answer in response to search request, as opposed to a document processing system which normally generates a list of references to a data base. It is a very specialized form of data management system in which the data structure contains items such as author name, title, publisher name, descriptive keywords, and possibly an abstract or full-text.

---

* CERTAIN COMMERCIAL SYSTEMS ARE IDENTIFIED IN THIS PAPER IN ORDER ADEQUATELY TO SPECIFY THE SYSTEMS BEING DESCRIBED. IN NO CASE DOES SUCH IDENTIFICATION IMPLY RECOMMENDATION OR ENDORSEMENT BY THE NATIONAL BUREAU OF STANDARDS, NOR DOES IT IMPLY THAT THESE SYSTEMS ARE NECESSARILY THE BEST FOR THE PURPOSE.

[1] Figures in brackets refer to the literature references on page 18.

The advent of time-sharing systems has made it possible to create automatic information handling systems that combine many of the services provided by standard library and documentation centers, with direct user participation in the search and retrieval process. Some systems also make it possible for the user to interact directly with the systems during the search and retrieval process. These on-line systems vary in their capability depending on the services provided and on the equipment available. Some systems, designed to be browsing tools, operate with the full text of documents displayed on a screen, while other systems store only bibliographic citations and possibly keywords.

Eight large-scale, operational or near-operational systems that are commercially available or government-owned were surveyed. Two systems were developed without a specific client. They are:

(1) The Mead Data Central developed by the Mead Data Corporation.

Note: This system was originally known as DATA CENTRAL.

(2) On-Line Retrieval Bibliographic Information Transfer (ORBIT) developed by System Development Corporation.

Six other systems were developed for a specific application and client. They are:

(3) The Central Information Reference & Control On-line (CIRCOL) developed by the Foreign Technology Division, Air Force Systems Command.

Note: The nucleus of this system is the Document Processor System (DPS) developed by IBM.

(4) Defence Documentation Center Information System (DDC) developed by the Defense Documentation Center.

(5) IBM Technical Information Retrieval Center (ITIRC) developed by IBM's Technical Information Retrieval Center.

(6) Medical Literature Analysis & Retrieval System (MEDLARS II) being developed by the Computer Science Corporation for the National Library of Medicine.

(7) The New York Times Information Bank (New York Times) being developed by IBM's Federal Systems Division for the New York Times.

(8) RECON/STIM developed by the Lockheed Missiles and Space Company for NASA.

Note: A nearly identical but proprietary version of this system is called DIALOG.

There are several experimental document processing systems operating with stored bibliographic citations. BOLD (Bibliographic On-Line Diaplay) [2] and TIP (Technical Information Project) [3] are examples. Another research project called INTREX (Information Transfer Experiment) [4] is currently under development at MIT. SMART (Salton's Magical Automatic Retrieval Technique) [5] is a fully automatic document processing system, capable of processing search requests in English and retrieving those documents most nearly similar to the search request. SMART can also be used for the evaluation of the effectiveness of different search methods. There are three on-line systems designed with emphasis on user orientation: AIM-TWX (Abridged Index Medicus-TWX) operated by the Lister Hill National Center for Biomedical Communication of the National Library of Medicine, BASIS - 70 (Battelle Automated Search Information System) developed at Battelles Columbus Laboratories, and SUNY (The State University of New York Biomedical Communication Network) developed by The State University of New York. No literatures on these systems exist except for some studies and plannings which preceded the actual documentation of the system. There are also bibliographic systems built by organizations for their own internal use. These systems are not included in this survey because they are not commercially available.

The purpose of this report is to present features of the systems in parallel fashion to facilitate comparison so that a potential user may have a basis for evaluation in terms of the capabilities which his requirements demand.

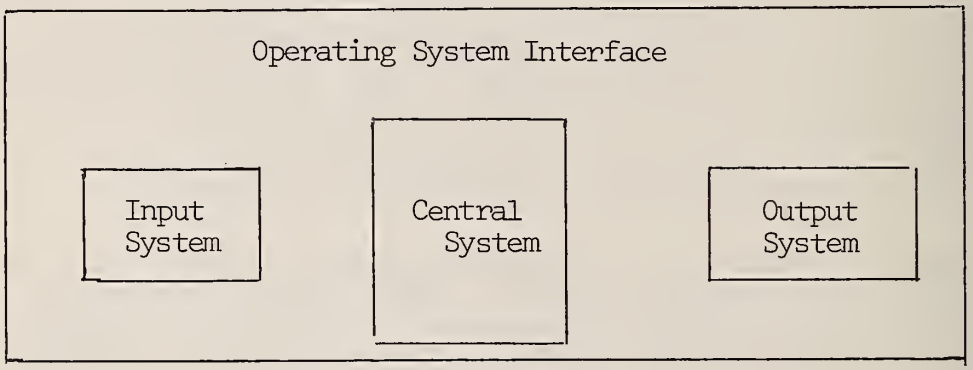
II. CLASSIFICATION OF A DOCUMENT PROCESSING SYSTEM

Document processing systems may be classified into two types in terms of their data base organization. First is the full-text type where the data base consists of the entire contents of the original documents; and second is the citation record type where the data base consists of formatted records containing author, title, descriptors, and other indices, and possibly some textual material. With the first type of organization, the document is readily available for browsing purposes, and every word is searchable; however, the space consumed is always much greater than in the second type. However, a full-text system could be set up as having retrievable segments, such as author, title, abstract number, etc. Not only all words in the abstract and title but all index terms are included on an inverted file. In this respect, the full-text system provides more flesibility in record structuring.

For the citation record organization the significant step is indexing against a vocabulary or a thesarus, because subsequent retrieval activities depend, to a large measure, upon the depth and accuracy of the indexing. Indexing is generally performed by a human who is specially trained in a particular subject area. Recently, much research effort has been done on automatic indexing by computer [6]. For the citation record type of organization, the full-text of the document is generally photographed and stored in microform for manual or mechanical retrieval. Any generalized data management system may be used for document processing with citation record type of organization. There is limited or sometimes no text processing capability and search terms are limited to only those that exists in the inverted file.

III. FUNCTIONAL COMPONENTS OF A DOCUMENT PROCESSING SYSTEM

A typical system can always be divided into three parts: the system input, the system itself, and the system output. The total system also has an operating system interface. A variety of ways of implementation exist as discussed below in the context of the systems surveyed.



A. Operating Environment

The operating environment consists of the specific computer system in which the document processing system will run. This includes hardware (the central processor plus the input/output devices and secondary storage devices) and software ("operating system" or "executive program" and interface program). The interface program is usually very much dependent on the mcahine configuration and the operating system.

B. System Inputs

The initial system input operation is the preparation of the data which will make up the data base. If the system is of the citation record type, then the next operation is the indexing of the documents and the creation of the citation records. Depending on the equipment available, some systems (e.g., New York Times) have on-line data entry with the indexer entering the data on a keyboard. The indexer uses a CRT to view the thesaurus or old documents in the system for cross-referencing purposes, and then constructs a record in a temporary work file. The above indexing procedure is sometimes called "machine-aided indexing" or "computer-assisted indexing". Other systems (e.g., CIRCOL) prepare the data input off-line and enter it into the system in the batch mode. A data definition language may exist (e.g., RECON/STIMS), enabling the system to be generalized for different applications.

At maintenance time, input in the form of update commands is needed. At the present time, the update is usually considered as a system function and the language is not user-oriented. A system analyst formulates the updates which are then usually run as a batch mode job. Some systems (e.g., MEAD Data Central and RECON/STIMS) allow updating in the background while searches are being conducted in the foreground. The user is cautioned against such practice since the file needed for searching may be "locked-up" while updating.

At retrieval time, input in the form of queries is entered into the system. For an on-line system, the query language is the major user language where simple yet powerful commands are stressed. A query language generally consists of commands made up of terms connected by Boolean operators and qualifiers.

A report generation time, output requirements are entered into the system. Most systems (e.g., CIRCOL and DDC) do not have a separate output report language but contain some options in the query language for specifying output requests. Some systems (e.g., Mead Data Central and ITIRC) provide user program linkages via code numbers whereby a user may write his own programs to format output reports.

C. The Central System

The major functions of the central system are to process user requests and to perform storage and retrieval of the data in the files. Factors such as file organization, search strategy, data accessing methods, type of peripheral equipment, internal representation of documents, sophistication of query language, etc., all affect the performance of the system. For tape systems (e.g., ITIRC), master records are organized sequentially. In order to facilitate a search, inverted files consisting of search words are set up.

ITIRC generates a separate tape file sorted according to word length. For the disk-oriented systems (e.g., Mead Data Central and CIRCOL), there exist dictionaries containing direct disk addresses. Mead Data Central maintains a range directory and a cascade type of search is conducted. CIRCOL's dictionary is sorted and a binary search is performed.

D. The System Outputs

The major functions of the system output are to prepare and display output reports. For most of the on-line systems, off-line outputs are available with the user specifying the output format. Some document processing systems (e.g., DDC, ITIRC and MEDLARS II) print out standard announcements or abstract bulletins at regular intervals. Some systems (e.g., RECON/STIMS) have a selective dissemination of information (S.D.I.) service by storing users' interest profiles, and the system outputs current items of relevant information within only those documents that match a user's interests. Some systems (e.g., RECON/STIMS) print out statistical information, for example the number of times a particular reference is retrieved.

IV. Overviews of the Systems

A prose description of each of the eight system surveyed is presented. Each description includes the identification of the system and its highlights. For detailed descriptions itemized under a feature list heading, the reader is referred to the Appendix II.

A. CIRCOL

CIRCOL (Central Information Reference and Control On-Line) exists as a specific implementation of a general teleprocessing - document processing system developed by the Foreign Technology Division, Air Force System Command. This system provides users with the capability to retrieve bibliographic and textual information from a large, user defined, computer stored data base. The CIRCOL data base is specifically designed to provide intelligence analysts with scientific and technical references of intelligence significance.

The CIRCOL system is a dynamic program structure consisting of three main modules: the system control program (PHENIX); the teleprocessing program (TP); and a modified version of IBM's 360 Document Processing System (DPS). DPS is a program package for processing unformatted textual information and runs under the IBM 360 Operating System (OS). TP, implemented under OS/360 release 18.6 with MVT, provides the on-line interface between DPS and the remote terminal user, and controls the execution of DPS.

The accumulation and processing of a data base query begins with TP which accepts search lines entered at a remote terminal and offers some acknowledgement of transmission to the terminal operator. TP uses these search lines to build an acceptable query for DPS. When the query has been completed, TP brings a copy of DPS into main storage and passes it to the query via the ATTACH feature of MVT. At this point, TP remains available to other terminals in the system while DPS gains control and interrogates the data base. Once the search has been evaluated, DPS returns control and any resulting output to TP (effectively removing itself from main storage) which then prints the resulting document information in a user controlled format on and/or off-line. DPS is not reenterable; however, up to three separate copies may be brought into main storage as needed so that three concurrent retrievals can be active. When the number of retrieval requests (completed queries) exceeds three at any one time, they are queried on a first-in first-out basis.

Error recovery procedures are provided by the system control program PHENIX which initiates and controls the execution of TP via the ATTACH feature of MVT. Under this system of varying levels of control, abnormal termination of DPS will not affect TP and abnormal termination of TP will not affect PHENIX. Thus, the system can be automatically restarted from the PHENIX level without human intervention.

B. DDC Information System

The Defense Documentation Center Information System may be regarded as an integrated system embodying several data bases. These data bases were developed since 1960 as parts of batch-oriented systems. DDC is developing an integrated on-line capability on the UNIVAC 1108 under EXEC 8. The data bases are:

Technical Report System (DD1473)
Work Unit Information System (DD1498)
Project Planning System (DD1634)
Contractor Performance & Evaluation System
Independent Research & Development System

The DDC On-line Information System utilize the above data bases. The prototype version is running. A major characteristic of the DDC On-line Information System is the front-end TPS (Text Processing System) as a data input. The TPS is interfaced via a Communication Terminal Module Control (CTMC) to EXEC 8 on the UNIVAC 1108 computer. Each CTMC unit will support up to 32 terminals. Another feature is the tutorial nature of the query languace. The computer guides the user at each step of query formation with a list of the available options.

This prototype On-line Information System is currently being evaluated. Internal expansion of the system to fully automate Agency operation is under consideration. Future developments may include integrated software for multi data banks, full-text system, machine-aided indexing, machine-generated theasurus and many others.

C. ITIRC

IBM's Technical Information Retrieval Center (ITIRC) operates an information retrieval system for searching normal text using a collection of programs called TEXT-PAC. TEXT-PAC consists of 30 programs written in Basic Assembly Language (BAL). The system requires an IBM/360 Model 40 or higher, using OS/360 MVT or MFT. Operation is in batch mode.

ITIRC has two major capabilities: the selective dissemination of information (IBM calls it current information selections (CIS)) and retrospective search. The source of inputs to the data file are engineering reports, patent applications, education materials, etc. These documents, after being coded and transcribed into machine-readable form, are entered into the computer. The machine does editing, formatting and proofing, and it outputs a text tape (for print purposes) and a search tape (sorted according to word length) for CIS and retrospective searching purposes. Besides these two tapes, there is also a third tape called OMAHA containing statistical information such as word frequency and spelling list.

CIS -- The ITIRC system provides subscribers, on a weekly basis, with selective notification of new data entering the system. The user fills out a CIS data sheet. Besides supplying some personal identifying data, he is encouraged to enter as many concepts as he thinks pertinent. The raw interest profile is converted to a machine-readable profile by a specialist and is entered into logical tables and processed against the search tape. Coincidences ('hits') are then sorted and collected and mailed to the user.

Retrospective Searching -- When the system user wants information from the complete file, an information retrieval specialist assists him by formulating queries to search the computerized file of abstracts. The retrospective search program selects those abstracts that match the search terms specified by the inquirer. The system output options allow selective printing of any paragraphs.

D. Mead Data Central

Mead Data Central is a generalized full-text information system developed by Mead Data Central, Incorporated. The system is capable of processing structured and unstructured data in an on-line conversational mode.

The main characteristic of this system is that it automatically takes every word (not on a "stop word" list which is predefined by the user) and arithmetic value in the file and places it in an inverted file in alphanumeric order, making it a searchable component of the data base. Associated with each component is a series of information strings such as relative position, security classification, and maintenance information. With this method, every word or value in the data base is searchable. There exists a "range directory" residing on direct access storage devices (DASD), and the component to be searched is first matched to obtain a pointer to the actual data location. The data themselves are in two forms. The serial file consists of a variable block length character string plus header information. The inverted file consists of the component followed by the associated information strings. Once the pointer is obtained for a query component, access is made to the DASD for sequential search.

The query language is used in a dialogue with the computer which allows dynamic modification of the query. It provides for concurrent search of both specifiable fields and free text. By virtue of the systems knowledge of word position in the sentence, the system also has a distance searching capability that makes it possible to search for the occurrence of two words within some number of words of each other.

Mead Data Central is the only system that provides the "KWIC-IT" option which uses colors to highlight the "hit" words of phrases. The Model CC-30 terminal provides four colors for output display. As used with the Mead Data Central System, an output record as displayed may show field designator in green. Successful matchings of selection criteria in red, ten significant words before and after the "hit" word or phrase in yellow, and all other information in blue. Such an output option facilitates browsing through the file.

E. MEDLARS II

The Medical Literature Analysis and Retrieval System (MEDLARS) is a mechanized bibliographic processing system. The first system, MEDLARS I, was developed by the General Electric Information Systems Division in 1964 and operated on a Honeywell 200-800 computer. It is a tape system. The system generates the monthly Index Medicus and the annual Cumulated Index Medicus for the National Library of Medicine. The system also performs demand searches. The second system, MEDLARS II, which is an improved version of MEDLARS I, is being designed by the Computer Science Corp.

MEDLARS II's detailed implementation is not yet final and little information is available at this time. NLM plans two versions of MEDLARS II, the initial system which will be available by the end of 1971, and the extended MEDLARS II. The main difference between initial MEDLARS II and the extended MEDLARS II is that the extended version will be an on-line system. Only the initial MEDLARS II is being reported on.

The initial MEDLARS II is implemented on IBM 360/50 with random-access disks. The data base is extended beyond that of MEDLARS I, MEDLARS II increases the capability in the areas of search parameters, bibliography, support of library functions, and it automatically maintains the vocabulary. One of the significant additions to the system is a data management module to facilitate handling of data and to provide a data description language which permits compilation to produce a table and a set of intermediate codes defining the file structures.

F. The New York Times Information Bank

The New York Times Information Bank, expected to be operational in the Spring 1971, will enable The New York Times to make its vast information files easily accessible to the general public.

The data base consists of abstracts and citations of articles in the New York Times and selected material fromover 60 other newspapers and periodicals. Actual clippings are mounted on paper and will be photographed at a reduction ratio of 25 to 1 and stored on 4" x 6" microfiche which will hold 99 images each. Within the New York Times building, the fiche will be stored in a Foto-Mem RISAR, a microfiche storage and retrieval device interfaced with the computer. The abstracts and citations will be entered by trained indexer-abstracters working at video terminals. The abstracts, terms, and other searchable elements will be entered into a temporary work file stored on disk. After the records are verified by a supervisor, a 'release' code will be applied and the records entered into the master file.

Inquirers will use video or typewriter terminals to enter queries consisting of descriptors connected by logical operators. The thesaurus and other user aids will be accessible to browsing via dialogue with the system. The outputs will be the abstracts of the documents with full citations, including the address of the associated clipping on microfiche. If the retrieval is within The New York Times Building, then the fiche may be viewed on the same terminal that was used for inquiries. Outside The New York Times, fiche storage, retrieval, and viewing will be manual. Master file updating is done every night in batch mode.

G. ORBIT II

ORBIT II (On-line Retrieval Bibliographic Information Transfer) is a bibliographic data storage and retrieval system developed by Systems Development Corp. which uses citations rather than full text. The system evolved from a batch system for intelligence purposes into an on-line, generalized system. There is a version called ORBIT II which operates under SDC's Time-Shared Executive program. The current version of ORBIT II operates under IBM OS/360.

The main characteristics of ORBIT II are its ability to handle very large files (more than 100,000 records) and to support a large number (more than 150) of on-line users concurrently. The system also has many tutorial features accessible via "EXPLAIN" and "?".

The package consists of two parts coded in PL/1: the file generation part, and the search and retrieval part. It is a proprietary system and little information is available on the internal file organization and the search strategies.

H. RECON/STIMS

The NASA information storage and retrieval process consists of two systems: RECON and STIMS. the RECON (REmote CONsole) system was developed by Lockheed for NASA to provide an on-line, conversational, retrieval access to the files produced and maintained by STIMS. STIMS (Scientific and Technical Information Modular System) was developed by Informatics TISCO for NASA to provide a batch processing file maintenance, search, and publications function.

The RECON/STIMS system is an information system capable of storing and retrieving scientific and technical documents. It runs on the IBM System 360 Model 50 or larger under OS/MFT II. The documents are manually indexed against the NASA thesaurus. These indexes are also tagged as being either of major or minor importance. When data enters into the batch input mode, the main file, which is called a linear file, and inverted files on indicated fields are constructed or updated. In the on-line mode it is only possible to post queries by using inverted index terms. However, in the batch mode one may search on any field in the record.

The system is also capable of doing SDI by limiting the search in a small accession number range or by generating a new inverted file for new documents and searching on it.

V. FEATURE LIST

There have been many attempts to develop a feature list which would characterize a generalized data management system e.g., [7], [8]. The same feature list would probably describe in part a document processing system. The purpose in developing the following feature list is to provide a checklist with short answers, thus avoiding long essay descriptions of each item. The feature list has the following major headings:

1. General Information - The non-technical details about the described system.

2. Operational Environment - The hardware configuation and the software restrictions imposed on the system.

3. Software Features - The facilities provided by the system.

4. User Interface - Various languages provided in order for the user to communicate with the system.

5. Internal Organization - The representation of information on a storage media.

6. Operational Functions - The functions and practices of the described system during execution.

The above major headings are further divided into sub-headings. There is no importance attached to the ordering. The following is the feature list headings and sub-headings with an explanation of each item.

1. GENERAL INFORMATION

1.1 System Name -- The name of the system in full as well as its acronym.

1.2 Source -- The name of the system originator or developer.

1.3 Plans for Maintenance & Improvement -- Planned extensions and type of maintenance to the system.

1.4 Type of Support -- The amount and type of supporting services provided by the system originator.

1.5 Availability -- Is the system in operation?

1.6 Cost -- The cost of the software if commercially sold or cost for hookup time if not sold.

1.7 User Population -- Names of organizations that are using the system.

1.8 Source Language -- The language in which the system is written.

1.9 Proprietary Software -- Is the software proprietary?

1.10 Documentation -- Any system manuals, operation manuals, or other formal documentation available on the system.

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (minimum configuration) -- This section consists mainly of the hardware configurations and the software restrictions imposed on the system.

2.1.1 Main Frame -- The name of the computer and its model number.

2.1.2 Input Devices -- cardreader, keyboard, etc.

2.1.3 Output Devices -- printer, CRT etc.

2.1.4 Mass Storage Devices -- tape, drum, disk, data cell, etc.

2.1.5 Document Storage Devices -- microfiche, microfilm, etc.

2.1.6 Communication Equipment -- teletypewriter, CRT, etc.

2.1.7 Core Size -- Minimum core memory size to run the system.

2.2 Operating System Version -- Name of the operating system.

2.2.1 Mode of Use -- batch or on-line, etc.

3. SOFTWARE FEATURES

3.1 Operating System Environment -- Any requirements on the operating system.

3.2 Transferability between Hardware -- Is it feasible to transfer the described system to other hardware?

3.3 Transferability between Operating Systems -- Is it feasible to transfer the described system to other operating systems?

3.4 Type of Security -- System security includes both the hardware security and software security via keys or passwords. Levels of security against accessing the data or against modifying the data are also mentioned.

3.5 Back-up Facility -- Whether the described system has a data back-up facility and if so, on what media. Back-up facility is sometimes provided by having a twin computer take over.

3.6 Restart & Recovery Capability -- The capability of the described system to recover and restart.

3.7 System Statistics -- Any form of statistical information that the described system is capable of generating.

3.8 Selective Dissemination of Inforamtion -- Whether the system has S.D.I. functions.

3.9 Indexing -- Does the system require indexing, and if so, what are the indexing procedures.

3.10 Thesaurus -- Whether the system has a thesaurus, and if so, what is the structure of the Thesaurus.

3.11 Input Data Editing and Validation -- The amount of checking performed on the input data.

3.12 Linkage to User's Code -- Whether the described system provides linkages such that user may write his particular application programs in assembly language, COBOL, FORTRAN, etc.

3.13 Special Feature -- Any special features that the system has.

4. USER INTERFACE

4.1 Data Description Language -- Whether the system allows the user to describe his own data.

4.2 Query Language -- Some highlights of the query language.

Devices -- Cardreader, teletypewriter, etc.

Language Type -- Procedural, near English, command type, etc.

Arithmetic Capability -- Whether arithmatic capability exists, and if so, what kind.

Boolean Logic for Selection -- Type of logical connectors.

Selection via Ranges of Values -- Type of arithmetic ranges and limits allowed.

Invocation of Predefined Queries -- Whether the queries may be saved and invoked at a later date.

Sample -- A sample of the query language, if available.

4.3 Output Report Language -- The mechanism for generating reports.

Device -- printer, teletypewriter, etc.

Language Type -- Procedural, same as query language, etc.

Prestored Format -- Is there the capability for storing frequently used output reports formats, and if so, how and when may such facilities be invoked.

On-line of Off-line Print Command -- For an on-line system, whether off-line printing is available.

Sort Specification -- At reporting time, whether sorting facilities are available.

Special Features Specification -- Any other features which may be specified at reporting time.

Sample -- A sample of the output report language, if available.

4.4 Maintenance & Update Language -- The procedure for updating.

Devices -- Cardreader, teletypewriter, etc .

Language Type -- Procedural, same as query language, etc.

Lockout Facility if On-line -- If updating is done on-line, the facilities for preventing simultaneous accesses of data.

Sample -- A sample of maintenance and update language, if available.

4.5 Browse Language -- Whether the full text or abstract is available to look over casually in order to select one to read.

5. INTERNAL ORGANIZATION

5.1 Data Base -- The logical nature of the files within the data base as the user sees it.

5.2 Data Structure -- The data as they are seen by the user. Does the data structure consist of hierarchical levels, repeating groups, fixed and/or variable length records, etc?

5.3 Storage Structure -- The organization of the data within a stored entry. Does the system maintain inverted lists, directories with pointers, etc?

6. OPERATIONAL FUNCTIONS -- The functions and practices of the described system during execution.

6.1 Data Access Method -- The way the stored data are accessed. It may be serial because the system uses tapes as its mass storage, or it may be random, because the system uses disk or drum. It may also be a combination of the above two methods.

6.2 Search Strategies -- The search strategies are related to the mass storage devices used and to the organization of the data. Any tricks or search optimizations are described here.

6.3 Update Facilities -- The update procedures and requirements imposed upon the software package by the practices of the system installation.

6.4 Time

6.4.1 Search Response -- If the system is on-line, the response time is critical. This item is difficult to assess since it is dependent on many factors, such as the way in which the system handles multi-programming, the number of terminals running simultaneously at that time, the size of the data base, the complexity of the queries, etc. An estimate is given if available.

6.4.2 Update Time -- This item is difficult to assess since it is usually dependent on the size of data base, and the amount of data to be updated. Also the time may increase if the update involves a major reorganization of the files. An estimate is given if available.

6.5 Space -- The amount of space devoted to the main file, inverted lists, and the ratio between the two. This item is very difficult to get because the size is usually growing so fast that even the system programmer in charge cannot keep track of it. Another factor is that the system may not be completely operational and therefore no studies have been made on this aspect.

VI. CONCLUSIONS

In this paper we have reviewed the characteristics of document processing systems. In addition, considerable attention has been paid to the description of a system via a feature list approach.

The state-of-the-art in on-line document processing systems has been moving very rapidly. The software progress in data base management, heuristic programming, automatic abstracting and indexing and also the hardware progress in front-end computers, optical character recognition devices, on-line data entry devices, etc., all have played a part.

The problem of system performance evaluation is not considered here because we still lack the tools in information science to determine precise performance measurements. Even if the desired measurements are hypothesized, there remains the interesting and difficult problem of quantifying system response. But I believe that this work has taken one step forward in analyzing a software product in terms of its component capabilities.

VII. ACKNOWLEDGEMENTS

I would like to express my appreciation to the following persons for devoting their time in reading and reviewing their respective sections:

The System Groups of ENMP for CIRCOL

Mr. J. Powers	for DDC
Mr. W. E. Brooks, Jr.	for ITIRC
Mr. M. Bayer	for Mead Data Central
Mr. J. Walden	for MEDLARS II
Mr. B. Falk	for the New York Times
Mr. C. Cuadra	for ORBIT
Mr. L. Stevens	for RECON/STIMS

I wish to thank Dr. Stephen J. Tauber, Mr. Daniel Minnick and Dr. Selden Stewart of NBS for reading earlier versions of this paper and for making valuable suggestions.

DISCLAIMER

All of these systems are changing, and this survey covers system capabilities up to the end of 1970. Every effort has been made to ensure the accuracy of the information contained in the system description. The writer assumes responsibility for any errors or misinterpretations which have entered the descriptions. Short answers to system features are given and the readers are requested to refer to the originator's source documents or manuals for more detailed descriptions.

VIII. REFERENCES

[1] Berul, L. H.: Document Retrieval, in Cuadra, C. (ed.) "Annual Review of Information Science and Technology," Vol. 4, Encyclopedia Britanica, Inc., Chicago, 1969.

[2] Burnaugh, H.P.: The BOLD (Bibliographic On-line Display) System, in Schecter, G. (ed.) "Information Retrieval -- A Critical View," Thompson Book Co., Washington, D.C., 1967.

[3] Mathews, W. D.: The TIP Retrieval System at MIT, in Schecter, G. (ed.) "Information Retrieval -- A Critical View" Thompson Book Co., Washington, D.C., 1967.

[4] Overhage, C. F. J.: Plans for Project Intrex, Scinece, Col. 152, No. 3725, (May, 1966).

[5] Salton, G.: Automatic Information Organization and Retrieval, McGraw-Hill, New York, 1968.

[6] Stevens, M. E.: Problems of Evaluation, in Automatic Indexing: A State-of-the-Art Report, NBS Monograph 91, Washington, D.C., March 30, 1965, 143-163.

[7] CODASYL System Committee: A Survey of Generalized Data Management Systems, CODASYL Systems Committee, technical report, May 1969.

[8] Fry, J. et.al.: Data Management System Survey, MTP-329, Mitre Corp., McLean, Virginia, Jan. 1969 (AD-684707).

APPENDIX I

Summary Chart

System Name	System Originator	Computer and Operating System	Full-text or Citation type	on-line or batch
CIRCOL	Foreign Technology Division, Air Force System Command, Wright-Patterson, Ohio	IBM 360/65 OS/MVT	Citation	on-line
DDC Information System	Defense Documentation Center, Cameron Station, Alexandria Virginia	UNIVAC 1108 EXEC 8	Citation	on-line (prototype)
ITIRC	IBM Technical Information Retrieval Center, White Plains New York	IBM 360/40 OS/MVT or MFT	full-text	batch
Mead Data Central	Mead Data Central, Inc.	IBM 360/40 DOS or OS	full-text	on-line
MEDLARS II	Computer Science Corporation for National Library of Medicine	IBM 360/50 OS/MVT	Citation	batch
New York Times Information Bank	IBM, Federal System Division for New York Times	IBM 360/50 DOS	Citation	batch
ORBIT	System Development Corportion, Santa Monica, California	IBM 360/40 OS/MVT or MFT	Citation	on-line
RECON/STIM	RECON written by Lockhead Missile & Space Co. and STIM written by Informatics TISCO, for NASA	IBM 360/50 OS/MFT II	Citation	on-line

APPENDIX II

DETAILED SYSTEM DESCRIPTIONS

The following are detailed notes on the document processing systems surveyed. Each of the systems is described in terms of the feature list presented above. Information not known to the writer is marked "unknown". The information was obtained through verbal briefings from the system representatives of that particular document processing system and from manuals, if available. Each section has been reviewed by the respective system representatives. All of these systems are changing, and this survey covers system capabilities up to the end of 1970. The writer assumes responsibility for any errors which have entered the descriptions; she would be pleased to be informed of corrections or additions.

CIRCOL

1. GENERAL INFORMATION

1.1 System Name -- CIRCOL (Central Information Reference and Control On-Line).

1.2 Source -- Foreign Technology Division, Air Force System Command, Wright-Patterson AFB, Ohio.

1.3 Plans for Improvement -- FTD plans to improve the overall CIRCOL system performance by taking every possible advantage of IBM 360 hardware/software advances.

1.4 Type of Support -- FTD consultation.

1.5 Availability -- Yes.

1.6 Cost -- Government owned and free to other government agencies.

1.7 User Population -- Air Force System Command Headquarters, Medical Intelligence Office, Harry Diamond Labs, Rome Air Development Center, Military Intelligence Agency (Redstone Arsenal), Defense Intelligence Agency, National Library of Medicine, Oceanographer of the Navy, Air Force System Command Divisions.

1.8 Source Language -- Assembly Language.

1.9 Proprietary Software -- No.

1.10 Documentation -- CIRCOL User's Guide, system documentation not complete.

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (minimum configuration)

2.1.1 Main Frame -- IBM 360 or 370 system which will support OS MVT.

2.1.2 Input Devices -- Teletypewriter, (IBM 2741 and IBM 2740). Data phone (AT&T or WU teletype models 33 and 35).

2.1.3 Output Devices -- Printer or terminals.

2.1.4 Mass Storage Devices -- Disk and/or data cells.

2.1.5 Document Storage Devices -- Microfiche (manually retrieved).

2.1.6 Communication Equipment -- Same as input devices.

2.1.7 Core Size -- 66K bytes for PHENIX - TP. 50K bytes for each copy of DPS.

2.2 Operating System Version -- 360 OS with MVT.

2.2.1 Mode of Use -- On-line query and batch file maintenance. Off-line retrieval of queries entered on-line.

3. SOFTWARE FEATURES

3.1 Operating System Environment -- IBM 360 Operating System with MVT.

3.2 Transferability between Hardware -- IBM 360 or 370.

3.3 Transferability between Operating Systems -- Within OS/360 MVT. (will run under release 19 or later MFT).

3.4 Type of Security -- Password associated with each terminal.

3.5 Back-up Facility -- Tape back-up of program and data base.

3.6 Restart & Recovery Capability -- Dynamic program structure allows for automatic restart of TP by PHENIX module. Searches in progress and partially accumulated queries are lost. DPS abnormal terminations mean only that the query in question cannot be evaluated, other users are not effected.

3.7 System Statistics -- User Activity Report (search times and number of documents retrieved).

3.8 Selective Dissemination of Information -- No, they hope to include this feature in the future.

3.9 Indexing -- Yes, indexing is computer-assisted with the system checking the input words against a controlled vocabulary.

3.10 Thesaurus -- There is no thesaurus, but the system has a controlled vocabulary file on disk. (Dictionary words can be listed).

3.11 Input Data Editing and Validation -- Yes, this is done in the preprocessor (Data Preparation Program).

3.12 Linkage to User Code -- No.

4. USER INTERFACE

4.1 Data Description Language -- The facility furnished with IBM's DPS is used - Data Base Description.

4.2 Query Language -- It is conversational consisting of question, and acknowledgment. The query is accomplished in six parts:

(1) Identification of the user.

(2) Identification of the application desired (DPS).

(3) Identification of data base desired (CIRCOL).

(4) Accumulation of a query.

(5) Qualification of the query, if desired.

(6) Specification of output.

The instructive nature of the system makes the query formation very easy with much interaction between the user and the system.

Device -- Teletypewriter.

Language Type -- Conversational with the system.

Arithmetic Capability -- None.

Logic for Selection -- Boolean operators exist for use as word connectors, while logical restrictors are available to define desired positional relationships of words in the document. In addition, users may further limit the acceptability of documents based on the bibliographic (fixed format) portion of the data. Fields within this portion may be examined using comparison operators.

Selection via Ranges of Values -- Yes, the comparison operator "BETWEEN" exists.

Sample -- See Figure 1, page 27

4.3 Output Report Language -- The language is part of the query language.

Devices -- Printer and teletypewriter.

Language Type -- Same as query language.

Prestored Format -- Format is defined at data base load time, but user may select certain options at output time.

On-line and/or Off-line Print Command -- Yes, the system asks the user whether he wants on-line and/or off-line output, and prints out accordingly.

Sort Specifications -- None.

4.4 Maintenance & Update Language -- The update is done via a modified batch DPS.

4.5 Browse Language -- No specific browse language.

5. INTERNAL ORGANIZATION

5.1 Data Base -- The CIRCOL data base is composed of three basic categories of foreign scientific and technical information presented in one fully integrated data base. These categories are: (1) Foreign Scientific and Technical Open Source Literature, (2) Intelligence Reports, and (3) Evaluated Intelligence Reports.

5.2 Data Structure -- Data structure consists of a formatted element called record or reference data and unformatted information called text. Although this information is data base dependent, CIRCOL record information includes: accession number, film number, type of document, date, country of information, and subject area. CIRCOL text information includes: descriptors, source, title, author, and in the more recently added documents, an abstract.

5.3 Storage Structure -- The data base consists of the following files:

(1) The Dictionary file is on a 2314 disk; records are sorted by alphabetic/numeric words. The remaining part of the record contains word frequency count, document frequency count, and a pointer to the Vocabulary file.

(2) The Vocabulary file is on a 2314 disk; each record contains a list of document numbers in which a particular word appears. This file, along with the Dictionary, serves as inverted files.

(3) The Master file contains all reference data (formatted) information and a coded form of the text data (unformatted). This file is directly accessed by the document number obtained from the Vocabulary file for checking relative keyword position and contents of formatted data fields. The Master file is the last file accessed during the search before retrieval from the Text file. The file storage device is a 2314 disk.

(4) The Text file contains the text portion of the document as it was entered into the data base. The Text file is directly accessed by the document number once it has been determined that the document satisfies the query. The storage device is a 2321 disk.

(5) Special files are built from terms whose number exceeds dictionary size limitations. These files enable searches to be made on such terms as though they were dictionary terms. The storage device is a 2314 disk.

6. OPERATIONAL FUNCTIONS

6.1 Data Access Method -- Direct access.

6.2 Search Strategies -- Binary search in dictionary file to obtain a pointer to the vocabulary file.

6.3 Update Facilities -- Update is done in the batch mode every two weeks with a separate program package. Words that cannot be found in the dictionary, may, optionally, be listed for manual analysis.

6.4 Time

6.4.1 Search Response Time -- The CIRCOL data base consists of approximately 500,000 documents. The search time averages 45 seconds.

6.4.2 Update Time -- It is batched. Time is a function of the amount of data to be updated.

6.5 Space -- The CIRCOL data base consists of approximately 600 million characters, 400 million of which make up the Text file.

Figure 1 - CIRCOL

SAMPLE SEARCHES

SIMPLE LONG FORM w/ONLINE REFERENCE

```
360 'CIRCOL' IN OPERATION
CIRCOL DATA BASE PRESENTLY CONTAINS REFERENCES TO APPROXIMATELY
400000 ARTICLES OR REPORTS FROM THE 1958-1969 TIME PERIOD
    ENTER TWO DIGIT STATION NUMBER
03
STATION 03 SIGNED ON
ENTER PASSWORD AND ROLL BACK PAPER BEFORE CARRIER RETURN (X-OFF).
XXXXXXXX PASSWORD OK.
DO YOU WANT LONG OR SHORT FORM OF CONVERSATION? L/S
l
    PLEASE IDENTIFY YOURSELF, LAST NAME FIRST
johnson
YOU ARE SIGNED ON USER JOHNSON
    ENTER YOUR SEARCH ONE LINE AT A TIME. LAST LINE MUST READ 'END'
AUTHORS MUST BE ENCLOSED IN @ SIGNS
SEE PAGE 12 OF CIRCOL USERS GUIDE FOR AN EXPLANATION OF AUTHOR RETRIEVAL
   1  OPTION CIRCOL  ,TEXT
   2  biologic & sabotage(+1)
   3  end
YOUR REQUEST IS BEING SERVICED BY DPS
       1 DOCUMENTS SATISFY YOUR REQUEST
    DO YOU WISH TO QUALIFY THIS REQUEST? Y/N
y
   ENTER QUALIFICATION STATEMENTS ONE LINE AT A TIME
LAST LINE MUST READ 'END'
SEE PAGE 17 OF CIRCOL USERS GUIDE FOR DEFINITIONS OF REFERENCE FIELDS
   6 if cntyussr eq y
   7 end
YOUR REQUEST IS BEING SERVICED BY DPS
       0 DOCUMENTS REMAIN
YOUR PREVIOUS SEARCH IS BEING REINSTATED
       1 DOCUMENTS SATISFY YOUR REQUEST
    DO YOU WISH TO QUALIFY THIS REQUEST? Y/N
n
    SPECIFY THE FORMAT OF YOUR OUTPUT BY LETTER
TO OBTAIN A LIST OF AVAILABLE OPTIONS, INPUT LIST
a
    DO YOU WANT ONLINE OUTPUT? Y/N
y
PLEASE STANDBY

297677     $03
ACCESSNR:  AP8033980
FILMNR:    1919/1691
    IF YOU WANT OFFLINE, ENTER AN 80 CHARACTER ADDRESS
IF NOT ENTER 'NONE'
none
    ARE YOU FINISHED? Y/N
y
YOUR JOB IS TERMINATED
PLEASE TURN OFF THE TERMINAL BEFORE LEAVING
CHOW
```

```
360 'CIRCOL' IN OPERATION
CIRCOL DATA BASE PRESENTLY CONTAINS REFERENCES TO APPROXIMATELY
400000 ARTICLES OR REPORTS FROM THE 1958-1969 TIME PERIOD
    ENTER TWO DIGIT STATION NUMBER
03
STATION 03 SIGNED ON
ENTER PASSWORD AND ROLL BACK PAPER BEFORE CARRIER RETURN (X-OFF).
XXXXXXXX PASSWORD OK.
DO YOU WANT LONG OR SHORT FORM OF CONVERSATION? L/S
s
    PLEASE IDENTIFY YOURSELF, LAST NAME FIRST
wilson
                    OK WILSON
BEGIN
   1  OPTION CIRCOL  ,TEXT
   2  hoodlum & helicopter(+1)
   3  end
TO DPS
      13 DOCS SATISFY
QUALIFY?
y
BEGIN
   6  if date gt 67
   7  and subjsode sc '01'
   8  end
TO DPS
       7 DOCUMENTS REMAIN
QUALIFY?
y
BEGIN
   8  and cntyussr eq y
   9  end
TO DPS
       6 DOCUMENTS REMAIN
QUALIFY?
y
BEGIN
   9  and classif lt 1
  10  end
TO DPS
       0 DOCUMENTS REMAIN
REINSTATING PREVIOUS
       6 DOCS SATISFY
QUALIFY?
n
SPECIFY OUTPUT FORMAT
OR 'LIST'
n
FINISHED?
y
THIS JOB TERMINATED
CHOW
```

BACKSPACE AND STRIKEOVER TO CORRECT

DDC Information System

1. GENERAL INFORMATION

1.1 System Name -- Defense Documentation Center (DDC) Information System.

1.2 Source -- Defense Documentation Center, Building 5, Cameron Station, Alexandria, Virginia 22314.

1.3 Plans for Maintenance & Improvement -- Extension of on-line capability within DDC for automation of duplicate checking, document identification, and reference inquiries. Conversion of batch retrieval applications to an on-line process. Extension of on-line capability externally to DoD Laboratories and other Federal agencies for direct access to technical and management information. Provide laboratories, commands, bureaus and ODDR&E with time-shared data management software for correlation and evaluation of information from several data bases, as well as the creation and maintenance of special files on-line.

1.4 Type of Support -- Defense Research and Development funds.

1.5 Availability -- DDC services are available to Defense activities, their contractors, and other Federal agencies.

1.6 Cost -- Nominal service charges are planned for the future.

1.7 User Population -- Defense research activities and their contractors primarily utilize DDC services. A limited on-line prototype system is being tested by NSA, Naval Ship Research and Development Center, the Air Force Weapons Laboratory, the Air Force Avionics Laboratory, the Air Force Materials Laboratory, Redstone Scientific Information Center, and one other site yet to be selected.

1.8 Source Language -- Sleuth (1108 Assembly language) and COBOL.

1.9 Proprietary Software -- No.

1.10 Documentation -- Available for review.

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (Minimum Configuration)

2.1.1 Main Frame -- UNIVAC 1108

2.1.2 Input Devices -- IBM 2741 terminals, and in the future CRT/keyboard devices.

2.1.3 Output Devices -- Pagewriter remote printers, highspeed impact printers, magnetic tape and COM units. Electrostatic printers in the future.

2.1.4 Mass Storage Devices -- Fastrand II drums, and disc systems in the future.

2.1.5 Document Storage Devices -- Microfiche, 16 and 35 mm roll film, now manually retrieved and reproduced for copy service. Future plans include possible use of automated full-text systems.

2.1.6 Communication Equipment -- Sixteen IBM Selectric 2741 terminals are used for data input. Nine Uniscope 300 CRT devices, and one KSR teletype terminal are used for retrieval and use of data management software for creation of special files. These are linked to the 1108 system via modems and a Communication Terminal Module Control (CTMC) unit that can service up to 32 terminals. Future plans include use of CRT terminals with tape cassettes and electrostatic printers for access to data input, retrieval, and data management software. Low cost teletype terminals will also be serviced.

2.1.7 Core Size -- 196,000 words, each word equivalent to 36 bits.

2.2 Operating System Version -- UNIVAC 1108 EXEC 8 (Level 25 and Level 26).

2.2.1 Mode of Use -- On-line query and batch file maintenance.

3. SOFTWARE FEATURES

3.1 Operating System Environment -- UNIVAC 1108 EXEC 8 real-time supervisory system.

3.2 Transferability between Hardware -- UNIVAC 1108 and to a limited extent, the 1107 system.

3.3 Transferability between Operating Systems -- only EXEC 8.

3.4 Type of Security -- The system is secure including hardware and software protection features.

3.5 Back-up Facility -- Now available through other govenment 1108 installations.

3.6 Restart & Recovery Capability -- A variety of file control and recovery procedures are employed.

3.7 System Statistics -- A wide variety of system statistics are available on equipment usage, products, and services.

3.8 Selective Dissemination of Information -- Both selective dissemination and demand services are available for obtaining copies of technical reports. The semi-monthly Technical Abstract Bulletin identifies recent document accessions. Current awareness services are also available to a limited number of DoD users.

3.9 Indexing -- Manual indexing using a thesaurus is current practice. Experiments using Machine-Aided Indexing are currently underway and appear promising.

3.10 Thesaurus -- A thesaurus is now used. Future plans provide for a machine-generated thesaurus based on actual terminology used.

3.11 Input Data Editing & Validation -- A series of edit checks are made on many data fields, including contract numbers, project numbers, and others.

3.12 Linkage to User Code -- No.

4. USER INTERFACE

4.1 Data Description Language -- A generalized file maintenance system is employed using decision edit tables for describing data fields and edit criteria.

4.2 Query Language -- The query language provides for tutorial assistance in use of the system on-line. A full range of Boolean search capabilities may be used as well as qualification search procedures for identifying only those records which meet given standards or limits.

Device -- Teletypewriter.

Language Type -- Conversational with the computer instructing the user of each option available.

Arithmetic Capability -- It can only sum a total of a set of fields.

Boolean Logic for Selection -- "AND", "OR", "NOT". Restriction: the "NOT" must not be the last condition in a query.

Selection Via Ranges of Values -- Range of dates may be specified.

Invocation of Predefined Queries -- Queries may be saved and invoked within the same run. Queries may not be saved after the user has terminated his run.

4.3 Output Report Language

Device -- Teletypewriter.

Language Type -- Same as query language.

Prestored Format -- There are four standard display formats at present. User may be able to specified parameters to the generalized report generator programs for any output format.

On-line or Off-line Print Command -- Yes.

Sort Specification -- Yes, fields may be specified for sorting.

4.4 Maintenance & Update Language -- Language is system programmer oriented.

4.5 Browsing Language -- No specific browse language.

5. INTERNAL ORGANIZATION

5.1 Data Base -- Several data bases are employed, each utilizing the same general logic of input edit, batch update, master file construction, and use with an inverted file for searching. Data files include the following:

Name	Function	Size
Technical Reports	Describes completed R&D	700,000 Records
Work Unit Information	Describes current R&D	40,000 Records
Project Planning	Describes future R&D	3,000 Records
Independent Research	Describes proposed R&D	6,000 Records
Contractor Performance	Describes quality of R&D	3,000 Records

5.2 Data Structures -- A record consists of header information followed by pointers to the relative position of each variable length field.

5.3 Storage Structure -- Master files are maintained on random access devices if used on-line, otherwise they are kept on magnetic tape. The inverted files are kept on the random access devices. The master files are organized by the control thesaurus.

6. OPERATIONAL FUNCTIONS

6.1 Data Access Method -- Direct access to the inverted file which is on Fastrand drum.

6.2 Search Strategies -- Unknown.

6.3 Update Facility -- Batched.

6.4 Time

6.4.1 Search Response -- Time to search is approximately 30-60 seconds depending on system load.

6.4.2 Update Time -- Time to update is a function of the data base size.

6.5 Space -- The master records occupy 23 reels of tape and the inverted files occupy approximately 1 to 2 reels of tape.

1. GENERAL INFORMATION

1.1 System Name -- ITIRC (IBM Technical Information Retrieval Center)

1.2 Source -- IBM, Technical Information Retrieval Center, White Plains, New York.

1.3 Plans for Improvement -- Unknown.

1.4 Type of Support -- TEXT-PAC, the nucleus of ITIRC, is a type 3 (IBM product, no support) package available through the Program Information Department.

1.5 Availability -- TEXT-PAC is available. ITIRC is not commercially available.

1.6 Cost -- Free.

1.7 User Population -- TEXT-PAC users consist of: Eastman Kodak, General Telephone and Electronics, and many others.

1.8 Source Language -- Basic Assembly Language of IBM 360.

1.9 Proprietary Software -- No.

1.10 Documentation -- 1. "Searching Normal Text for Information Retrieval" IBM, Data Processing Application, White Plains, New York 10601. 2. "TEXT-PAC Basic Documentation" available through IBM, Program Information Department.

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (Minimum configuration)

2.1.1 Main Frame -- IBM 360/40

2.1.2 Input Devices -- Card reader, tapes.

2.1.3 Output Devices -- Printer, tapes.

2.1.4 Mass Storage Devices -- 2311 disk and 4 nine-track tapes.

2.1.5 Document Storage Devices -- Tapes.

2.1.6 Communication Equipment -- 7711 Tape transmission unit.

2.1.7 Core Size -- 256k with 128k region available for the program.

2.2 Operating System Version -- 360 OS with MVT or MFT.

2.2.1 Mode of Use -- Batch.

3. SOFTWARE FEATURES

3.1 Operating System Environment -- IBM OS/360 MVT or MFT.

3.2 Transferability between Hardware -- Within IBM 360.

3.3 Transferability between Operating Systems -- Within IBM OS/360.

3.4 Type of Security -- 1 level of data access security, either yes or no. Data modification is not allowed.

3.5 Back-up Facility -- The text tape also serves as a data back-up tape. There is no computer back-up facility mentioned.

3.6 Restart & Recovery Capability -- Yes.

3.7 System Statistics -- Forms are distributed to users to get continuous feedback on their satisfaction with the performance of the system.

3.8 Selective Dissemination of Information -- Yes, user fills out a data sheet consisting of personal data, job data and special search words applicable to his needs. The IR specialist takes the user's data sheets and creates a profile similar to the query language form. This profile is stored in the system. The incoming document is processed against the stored profiles. The notification and response card provided is a special double-card. The left hand card contains the bibliographic data and abstract for each answer. The right-hand card is used first to ask the user to make an appropriate response in regard to his profile and, second, it is used to order the complete document or a microfiche copy.

3.9 Indexing -- No, because it is a full-text system.

3.10 Thesaurus -- No.

3.11 Input Data Editing & Validation -- Yes, there is even a spelling check.

3.12 Linkage to User's Code -- User may write his own output report format in assembly language and link the code to ITIRC by using the print control code which is a number from 000 to 999.

4. USER INTERFACE

4.1 Data Description Language -- No data description language is available. Documents entering the system are assigned a printed control code, such as title = 000, author = 200, etc., up to 999.

4.2 Query Language -- User supplies the query to the Information Retrieval Specialist who formulates the query, keypunches and batches it for the daily search run against tapes. Answers are in the mail the next day.

Device -- Cardreader.

Language Type -- Stylized English-like language. Word stem may be used by allowing "$" to appear at the place where stemming may occur.

Arithmetic Capability -- None.

Boolean Logic for Selection -- "AND", "OR", and "NOT".

Selection via Ranges of Values -- None.

Invocation of Predefined Queries -- Yes, the interest profiles are predefined queries.

Sample --

A1	ON ADJ LINE
A2	ONLINE OR ON-LINE
A3	REAL ADJ TIME
A4	A1 OR A2 OR A3
A5	INFORMATION ADJ SYSTEM OR RETRIEVAL OR SERVICE
CONC	A4 and A5

4.3 Output Report Language -- There is a standard output on a periodic schedule. In addition, a Key Word Out-Of-Context (KWOC) is pre-

pared. The demand output is described as follows:

Device -- Cardreader.

Language Type -- System oriented in the form of a programming language.

Prestored Format -- There are 999 print-control codes which may be used for formatting the input and output. User requests paragraphs he wishes to see.

On-line or Off-line Print Command -- It is not an on-line system.

4.4 Maintenance and Update Language -- To correct a word in the text, the relative position of the word has to be known. The correction, plus the document number, paragraph number, line number, and word number must be indicated. All corrections are processed against the original tape and outputs a corrected edit tape.

4.5 Browse Language -- No browsing capability in this version.

5. INTERNAL ORGANIZATION

5.1 Data Base -- The permanent file is on three tapes: text tape, search tape, and OMAHA tape. There are many files in the data base: IBM, NON-IBM, JOURNAL and INVENTION, etc.

5.2 Data Structure -- In the search tape, records are organized by word length with pointers indicating the start of the word groupings. In the text tape, records are organized by print control characters for each paragraph.

5.3 Storage Structure -- Tape oriented serial records.

6. OPERATIONAL FUNCTIONS

6.1 Data Access Methods -- Tape oriented serial record. Within record, words are sorted into groups by the number of characters.

6.2 Search Strategies -- Serial from record to record, but within a record words are searched only in the specified word length group.

6.3 Update Facilities -- Update weekly from the forms.

6.4 Time --

6.4.1 Search Response -- 1 to 2,000 documents/minute.

6.4.2 Update Time -- Unknown.

6.5 Space -- Unlimited tape space.

MEAD DATA CENTRAL

1. GENERAL INFORMATION

1.1 System Name -- Mead Data Central, formerly known as Data Central.

1.2 Source -- Mead Data Central, Inc. (MDCI).

1.3 Plans for Improvement -- Extensive on-going effort.

1.4 Type of Support -- Complete requirement analysis, data conversion, programming support, etc. at MDCI Service Center.

1.5 Availability -- Through MDCI Service Centers since 1968.

1.6 Cost -- Published rate schedule.

1.7 User Population -- Environmental Protection Agency, National Aeronautics Space Administration, Health, Education & Welfare, Department of Defense, National Institute of Health, United States Air Force, National Technical Information Service, Union Carbide, New York and Ohio Bar Associations, American Psychological Association, Corporation for Research in Social Sciences (CRESS).

1.8 Source Language -- Assembly, COBOL, FORTRAN.

1.9 Proprietary Software -- Yes.

1.10 Documentation -- On request for specific user requirements.

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (minimum configuration)

2.1.1 Main Frame -- IBM 360/40 and up.

2.1.2 Input Devices -- Data input on IBM Magnetic Tape/Selectric Typewriters (MTST); on-line remote terminals, especially CRT's. Query input on on-line remote terminals (especially CRT's).

2.1.3 Output Devices -- Same as Query Input devices.

2.1.4 Mass Storage Devices -- Direct Access Storage Devices (DASD).

2.1.5 Document Storage Devices -- DASD.

2.1.6 Communication Equipment -- Primarily color CRT terminals and various other remote terminal devices.

2.1.7 Core Size -- Variable depending upon Operating System and number and type of communication lines supported. For example, under DOS and supporting ten half-duplex, dial-up communication lines, the core requirement is 64K.

2.2 Operating System Version -- IBM 360 DOS or OS.

2.2.1 Mode of Use -- Time-shared. The foreground partition used for queries and the background partition for file updating.

3. SOFTWARE FEATURES

3.1 Operating System Environment -- IBM 360 DOS or OS.

3.2 Transferability Within Hardware -- Within IBM 360 or 370 family.

3.3 Transferability Within Operating System -- Within 360 DOS or OS.

3.4 Type of Security -- User security code may be changed daily. Each entry and/or field (segment) may be given a security code.

3.5 Back-up Facility -- Additional MDCI Service Centers. Data bank back-up is available in magnetic form.

3.6 Restart and Recovery Capability -- Unknown.

3.7 System Statistics -- Unknown.

3.8 Selective Dissemination of Information -- Unknown.

3.9 Indexing -- It is a full-text system and therefore, indexing is not required.

3.10 Thesaurus -- Uncontrolled, computer-generated, available per database.

3.11 Input Data Editing and Validation -- Yes, per data owner specifications.

3.12 Sorting -- At output reporting time the system will ask the user whether he wants the entries sorted, if yes, enter the name of the field(s) to be sorted.

3.13 Special Features -- This system is capable of automatically restructuring the file without rebuilding. Multi-file or cross-file searching is also available. The system is also capable of doing "recursive search" meaning using the previously obtained answers as input terms to the next query. There is a "superfield concatenation" capability in which the user may concatenate fields and define a super-field for searching purposes.

4. USER INTERFACE

4.1 Data Description Language -- The user provides MDCI with input data specifications and data base specifications. MDCI will use these to set up programs in assembly language, COBOL, or FORTRAN called Data Base Definition and Input Data Definition.

4.2 Query Language --

Device -- See query input device (2.1.2).

Language Level -- English-like dialogue.

Arithmetic Capability -- Yes.

Boolean Logic for Selection -- Major "AND", minor "AND", "OR".

Selection via Range of Values -- Yes.

Sample -- See Figure 2, page 43

4.3 Output Report Language -- The output format is programmable via Assembly, COBOL, or FORTRAN language and stored per key-name. Names or numbers are used to specify the pre-stored format. Device may be specified by "console" or "printer". Computer will also offer (to hard copy devices) a chance to roll the paper ahead.

4.4 Maintenance and Update Language -- Unknown.

4.5 Browse Language -- Full flexibility available.

5. INTERNAL ORGANIZATION

5.1 Data Base -- The data base consists of three main files: the serial file, the inverted search index file, and the range directory.

5.2 Data Structure -- Variable per owner specifications. Up to 61,441 segments per entry and up to 255 files per data base may be defined.

5.3 Storage Structure -- The records are organized randomly on the DASD which is pointed to from the range directory. The new data is added or inserted at the end with pointers in the range directory pointing to it. There is an inverted list maintained and every word or value is inverted except those in a common "stop-word" list.

6. OPERATIONAL FUNCTIONS

6.1 Data Access Methods -- Proprietary special index-sequential access method.

6.2 Search Strategies -- Words are searched in the directory to find the proper range and then sequential search within the range.

6.3 Update Facilities -- Batch mode update.

6.4 Time

6.4.1 Search Response -- In minute(s) dependent upon search complexity.

6.4.2 Conversation & Browsing Response -- In seconds.

6.4.3 Update Time -- Dependent upon data base size and amount of data to be updated.

6.5 Space -- The inverted files space is dependent upon the original file. It occupies about 20 to 60 percent of the original file. The average is approximately 35%.

Computer: YOU ARE NOW IN COMMUNICATION WITH (DATA) CENTRAL.
PLEASE ENTER 10 CHARACTER IDENTIFICATION.

User: 1234567890

Computer: ENTER FILE, MESSAGE OPTION

User: projects, long

Computer: ENTER REQUEST

User: mercury

Computer: THERE ARE 9 ENTRIES THAT SATISFY YOUR REQUEST.
DO YOU WANT TO PROCESS ANSWERS: NO, PRINT, MODIFY?

User: modify

Computer: ADD NUMBER 002 MODIFICATION

User: and $startdate gtr Oljan69

Computer: THERE ARE 7 ENTRIES THAT SATISFY YOUR REQUEST.
DO YOU WANT TO PROCESS ANSWERS: NO, PRINT, MODIFY?

User: print

Computer: ENTER DESIRED OUTPUT, DEVICE

User: full-ret, console

Computer: DO YOU WANT THE ENTRIES SEQUENCED BY ANY OF THE RETRIEVED SEGMENTS? YES OR NO.

User no

1. GENERAL INFORMATION

1.1 System Name -- MEDLARS II (Medical Literature Analysis and Retrieval System) initial version.

1.2 Source -- Software written by Computer Sciences Corporation for National Library of Medicine.

1.3 Plans for Improvement -- An extended system which is on-line is being planned.

1.4 Type of Support -- National Library of Medicine will maintain.

1.5 Availability -- Initial system is expected to be operational by the end of 1971.

1.6 Cost -- At present, the system is not intended to be commercially available.

1.7 User Population -- The National Library of Medicine.

1.8 Source Language -- PL/1 and ALC.

1.9 Proprietary Software -- No.

1.10 Documentation -- The Principles of MEDLARS, National Library of Medicine, (no date).

There are several internal documents, but not publicly available.

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (minimum configuration)

2.1.1 Main Frame -- IBM 360/50

2.1.2 Input Devices -- Keymatic (keyboard & magnetic tape), card reader.

2.1.3 Output Devices -- Printer or tape for photo-composition.

2.1.4 Mass Storage Devices -- Magnetic tape, 2314 disk.

2.1.5 Document Storage Devices -- Not part of the MEDLARS II system.

2.1.6 Communication Equipments -- Not on-line.

2.1.7 Core Size -- 512K main memory and 1 million LCS. (There is version for demand searches only which requires 256K with no LCS).

2.2 Operating System Version -- 360 OS/MVT (Demand searches version will run under OS/MFT).

2.2.1 Mode of Use -- Batch only, program re-entrant.

3. SOFTWARE FEATURES

3.1 Operating System Environment -- 360 OS/MVT. It operates under an interface control program called COSMIS.

3.2 Transferability between Hardware -- Within IBM 360 and 370.

3.3 Transferability between Operating System -- OS/MVT.

3.4 Type of Security -- Security for updating files is available, but no security at present is provided for accessing the files.

3.5 Back-up Facility -- Tape back-up.

3.6 Restart & Recovery Capability -- Checkpoint restart is available for long runs.

3.7 Selective Dissemination of Information -- No S. D. I. based on interest profile, but the system periodically generates standard outputs called Index Medicus, Cumulated Index Medicus, recurring bibliographies and literature searches on specific topics.

3.9 Indexing -- Manual.

3.10 Thesaurus -- Yes.

3.11 Input Data Editing & Validation -- Yes.

3.12 Sorting -- No.

4. USER INTERFACE

4.1 Data Description Language -- The data description language is compiled by an ALC program. The output from this compiler is a set of data description tables which define the file structure and modules of DMOPS interpretive programs. DMOPS is a 'machine independent' object code which is interpreted by the interpreter. The data description language provides the ability to build directories or inverted files on any number of items. The language is comprised of four kinds of statements: FILE, RECORD, field description, END. The language is built upon keys, and reserved word descriptors. Each descriptor begins with a clause which may contain other key words.

Sample -- See Figure 3, page 49

4.2 Query Language -- There are two types of query formation. The user may use the LPS (Library Processing System) language or he may fill out forms designed for search formation. There is a "Form Preprocessor" which will take the form entry and convert it into the language. The retrieval consists of either a key request which will cause a unique item to be retrieved from the system; or a query request which is a boolean expression which will yield a collection of items covering a limited range of interests.

Device -- Key punched or tape.

Language Type -- Forms or language delimited by reserved words.

Arithmetic Capability -- None.

Boolean Logic for Selection -- "AND", "OR"

Selection Via Ranges of Values -- Yes, one can specify a "limit".

4.3 Output Report Language -- There are also forms designed for output report specification. The default is standard format.

Device -- Key punched or tape.

Language Type -- Form specification.

Prestored Format -- Yes.

On-line or off-line Print Command -- Not applicable because the initial MEDLARS II is not on-line.

Partial printout -- Yes.

4.4 Maintenance & Update Language -- It is also done by filling a form. The "form proecessor" will generate a language which is command oriented, with a command name (such as ADD, DELETE, UPDATE, REPLACE) followed by a list of parameters in parenthesis.

4.5 Browse Language -- None.

5. INTERNAL ORGANIZATION

5.1 Data Base -- There are four data bases in MEDLARS II.

1. Item record data base - A record for every journal title in the library.

2. Augmented MeSH - A collection of valid indexing terms plus scope notes.

3. Citation record data base - A collection of citations supported by the Augmented MeSH file.

4. Supporting data base - A collection of query formation, system statistic and management statistic package, and all other housekeeping functions.

5.2 Data Structure

File Format -- The file contains citation records segmented and dynamically allocated.

Record Format -- The record consists of a fixed part required, fixed part not-required, variable part required, variable part not-required. Hieriachial structure is allowed.

5.3 Storage Structure

Secondary Storage Organization -- There is an available space table to assign space on a track. A record locator file accessed via an accession number contains the relative track address on the disk.

Inverted List Maintained -- User option to define inverted files.

6. OPERATIONAL FUNCTIONS

6.1 Data Access Methods -- Absolute address of disk is obtained and directly accessed.

6.2 Search Strategies -- The terms of the search equation are analyzed. Search is performed on the inverted file, and then linear search on the subsets.

6.3 Update Facilities -- Batched.

6.4 Time -- Unknown, because the system is not yet operational.

6.5 Space -- The inverted files occupy approximately 25% as much space as the original file.

AN EXAMPLE OF DATA DEFINITION LANGUAGE OF MEDLARS

```
FILE PAYROLL:  DATA BASE;
     RECORD PERSONNEL:  REQUIRES (EMP-NO, DEPT, RATE);
       EMP-NO:  DECIMAL, SIZE IS 8 BYTES, DIRECTORY UNIQUE;
       EMP-NAME:  REQUIRES LAST, CONTAINS (FIRST, MIDDLE);
           FIRST:  CHARACTER, SIZE IS VARIABLE BYTES;
           MIDDLE:  CHARACTER, SIZE IS 1 BYTE;
       DEPT:  DECIMAL, SIZE IS 6 BYTES;
       RATE:  DECIMAL, SIZE IS 4 BYTES;
       WORK CAT:  BINARY, SIZE IS 8 BITS, ALLOW (7=3 THRU 2=500);
       REPORT-TYPE:  BINARY,  SIZE IS 16 BITS, DIRECTORY COORDINATE
                     FORMAT-CAT;
       FORMAT-CAT:  BINARY, SIZE IS 16 BITS, DIRECTORY COORDINATE
                    REPORT-TYPE;
       REPORT GROUP:  CONTAINS (REPORT-TYPE, FORMAT-CAT), OCCURS AS
            SHOWN;
END PAYROLL;
```

FIGURE 3 Sample DDL of MEDLARS

NEW YORK TIMES INFORMATION BANK

1. GENERAL INFORMATION

1.1 System Name -- The New York Times Information Bank

1.2 Source -- All software not written by Times staff is written under contract by IBM, Federal System Division, Gaithersburg, Maryland.

1.3 Plans for Improvement -- Times staff with some IBM contractual arrangement will maintain and improve the system.

1.4 Type of Support -- The rights of the program belong to the New York Times. They will consider software leasing at a presently undefined cost.

1.5 Availability -- First half of 1971.

1.6 Cost -- Unknown.

1.7 User Population -- The New York Times and outside subscribers.

1.8 Source Language -- Basic Assembly language and PL/1.

1.9 Proprietary Software -- Yes, New York Times solely owns all the software.

1.10 Documentation -- Unknown.

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (Minimum configuration)

2.1.1 Main Frame -- IBM 360/50.

2.1.2 Input Devices -- In-house terminals (used for data entry, inquiry and output) are IBM 4506 display stations with IBM 4279 terminal control units.

2.1.3 Output Devices -- Video terminals as in 2.1.2, IBM 1403 high speed printer with upper and lower case.

2.1.4 Mass Storage Devices -- IBM 2314 disk, IBM 2321 data cell, IBM 2401 tape drives.

2.1.5 Document Storage Devices -- Foto-Mem RISAR (4.95 million page images) controlled by a CENTAUR computer.

2.1.6 Communications Equipment -- see Input Devices as 2.1.2.

2.1.7 Core Size -- 512k, but the system only uses 200k bytes.

2.2 Operating System Version -- IBM 360 DOS.

2.2.1 Mode of Use -- On-line query and batch file maintenance.

3. SOFTWARE FEATURES

3.1 Operating System Environment -- IBM 360 DOS partition controlled task.

3.2 Transferability between Hardware -- Within IBM 360.

3.3 Transferability between Operating System -- Within 360/DOS.

3.4 Type of Security -- Data access security is available via customer assigned identification number and password. Data modification is not allowed.

3.5 Back-up Facility -- Unknown.

3.6 Restart & Recovery Capability -- Yes.

3.7 System Statistics -- Yes.

3.8 Selective Dissemination of Information -- Not planned at present but the capability exists in the system.

3.9 Indexing -- Yes, indexing is computer-assisted with the system checking for valid words against the thesaurus.

3.10 Thesaurus -- Yes.

3.11 Input Data Editing and Validation -- Yes, both on-line and off-line.

3.12 Linkage to User Code -- None.

4. USER INTERFACE

4.1 Data Description Language -- None.

4.2 Query Language -- Interactive dialogue with the system.

Device -- Terminals.

Language Level -- User oriented dialogue with the system.

Arithmetic Capability -- None.

Boolean Logic for Selection -- 'AND', 'OR', 'NOT'.

Selection via Ranges of Values -- Dates, sources, descriptor and abstract weights, etc.

Invocation of Predefined Queries -- No.

4.3 Output Report Language -- Abstracts, citations and microfiche addresses are outputted via the terminal or off-line. Hard copy of abstracts and full text may be obtained on request.

4.4 Maintenance & Update Language -- Stylized format to be used only by system programmer.

4.5 Browse Language -- Yes, the computer guides the user by asking at each step whether the user would like to see the abstract.

5. INTERNAL ORGANIZATION

5.1 Data Base -- The data base consists of three files. The descriptor file on disk, the locator file on disk and the abstract file on data cell.

5.2 Data Structure -- No information is given but some items are mentioned in each file. The descriptor file contains the terms, term type, searchable title (from a list of 200), and time period. The locator file contains bibliographic information such as by or about a man, the source (N.Y. Times, other journals, wire services, etc.), types of materials (letters to editor, editorial, etc.). The abstract file contains the detailed abstract of the document in text form and is only retrieved when all the search criteria have been met.

5.3 Storage Structure -- Proprietary information.

6. OPERATIONAL FUNCTIONS -- This system is not operational, the software is proprietary and therefore, no information is given.

ORBIT II

1. GENERAL INFORMATION

1.1 System Name -- ORBIT II (On-line Retrieval Bibliographic Information Transfer).

1.2 Source -- System Development Corporation, 2500 Colorado Avenue, Santa Monica, California 90406.

1.3 Plans for Improvement -- SDC plans to improve the system such that it will handle several different data bases with one copy of the Retrieval Program in one partition.

1.4 Type of Support -- SDC will maintain the system for one year. After the first year SDC will continue to provide maintenance on the basis of separate contract.

1.5 Availability -- Available in January 1971.

1.6 Cost -- $22,000 if purchased, or the system may be leased at $1,200 plus a monthly charge of $750, which reduces to $600 per month after 12 months.

1.7 User Population -- Unknown.

1.8 Source Language -- PL/1.

1.9 Proprietary Software -- Yes.

1.10 Documentation -- Users and Operator Manuals.

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (Minimum configuration)

2.1.1 Main Frame -- IBM 360/40.

2.1.2 Input Devices -- Standard phone-coupled terminals, such as teletype, IBM 2741, Time sharing terminal 707 Execuport, or Vernitron. Also CRT terminals such as CC-335 or Datapoint 3300.

2.1.3 Output Devices -- Teletypewriter, off-line printer.

2.1.4 Mass Storage Device -- 2314 disk.

2.1.5 Document Storage Device -- None.

2.1.6 Communication Equipment -- IBM 2701 Data Adapter Unit (for up to about 8 ports). IBM 2702 Transmission Control (for about 24 to 32 ports). IBM 2703 Data Transmission Control (for up to 96 ports). A special IBM teleprocessing procedure, called QTAM (Queued Telecommunications Access Method) must be used with the above equipment to handle the incoming and outgoing messages to the system.

2.1.7 Core Size -- Minimum 256k bytes of core storage: 160k for ORBIT II, 40k for QTAM, and 24k for OS/MFT.

2.2 Operating System Version -- IBM 360/OS/MFT or IBM 360/OS/MVT.

2.2.1 Mode of Use -- On-line and batch.

3. SOFTWARE FEATURES

3.1 Operating System Environment -- OS/MFT, OS/MVT

3.2 Transferability between Hardware -- Within IBM 360 and IBM 370.

3.3 Transferability between Operating System -- OS/MFT, OS/MVT.

3.4 Type of Security -- Optional.

3.5 Back-up Facility -- If desired.

3.6 Restart and Recovery Capability -- Yes.

3.7 System Statistics -- Unknown.

3.8 Selective Dissemination of Information -- Limited.

3.9 Indexing -- Manual. However, an "automatic indexing" feature could be added with the addition of about $2,500.

3.10 Thesaurus -- Not part of the standard package.

3.11 Input Data Editing & Validation -- Yes.

3.12 Sorting -- No.

4. USER INTERFACE

4.1 Data Description Language -- There is no data description language. However, the user will have to provide SDC with specifications for the data base. SDC will use these to prepare a file structure description deck (specific to each data base) and provide the customer with the file generation program.

4.2 Query Language -- Easy to use with a lot of tutorial and detail error diagnostics. The commands may be entered in any sequence.

Device -- Teletypewriter 2741 and CRT.

Language Type -- User oriented.

Arithmetic Capability -- None.

Boolean Logic for Selection -- Unrestricted use of all boolean.

Selection via Ranges of Values -- Search for term adjacent alphabetically up and down.

Sample -- See Figure 4, page 59 and 60.

4.3 Output Report Language

Device -- Teletypewriter 2741.

Language Type -- Same as Query Language.

Prestored Format -- Yes.

Sort Specification -- Yes, the system provides for ordering the outputs in terms of relevance or any one of several other numeric categories.

Off-line or On-line Print Command -- Yes.

Special Features Specification -- Specification may be given to print only parts of the record.

4.4 Maintenance & Update Language -- On-line update is possible. The language is the same as the query language and consists of commands followed by specifications.

4.5 Browse Language -- There is no browsing capability of the original documents because this is not a full-text system. However, the indexed terms may be browsed by the use of "NEIGHBOR", "ROLL-DOWN" , and "ROLL-UP" commands.

5. INTERNAL ORGANIZATION

5.1 Data Base -- There are three main files: Unit record file, Postings file, and Locator file.

5.2 Data Structure -- The data structure consists of category name (such as author, title, indexing terms, etc.) and data value in alphanumeric and special symbols. ORBIT II can handle up to 255 on a unit record. Hierarchical data structure is available.

5.3 Storage Structure -- Unknown because it is a proprietary software package.

6. OPERATIONAL FUNCTIONS

6.1 Data Access Method -- Unknown, it is a proprietary software package.

6.2 Search Strategies -- Unknown, it is a proprietary software package.

6.3 Update Facilities -- Both on-line and batch update facilities are provided. There is a limit as to how much on-line updating can be done before the data base needs to be reconstructed. The user may determine the amount of space to be left in the file by the File Generation Program for on-line additions.

6.4 Time -- There is no time quoted for response to a query. The following are times quoted for batch-mode updating on a 360/67:

- Building an original file of 3,000 records requires 5 minutes of batch time.
- Building an original file of 30,000 records requires 2 hours of of batch time.
- Adding 30,000 records to 60,000 record data base requires 2 hours of batch time.
- Adding 3,000 records to a 130,000 record data base requires 40 minutes of batch time.

6.5 Space -- The space required on the IBM 2314 disk is approximately equal to the number of characters in the main data base, plus 50% of that number for the special index files.

ANNOTATIONS

PRINTOUT

NOTE:
USER INPUT IN LOWER CASE, MACHINE OUTPUT IN UPPER CASE

The HELLO message, available as an option, has various uses, depending upon customer needs. Here it provides status information and a security key.

The user is alerted that it is his turn by the cue word "USER:". (He is also informed about any error he makes.)

If the user so desires, the program will provide additional information about its operation.

The user enters an EXPLAIN command (used to obtain explanations of any command or program message). Commands are surrounded by quotation marks.

The number of elements and those which are searchable are separately determined for each data base. However, all are printable.

The user is asked to enter a search term or command. Note automatic numbering of search statements.

Any term of any length including numeric terms may be entered on a search term. The user enters air pollution as a search term and is informed that 423 records in the data base are indexed under that term. Had there been no postings, he would have been so advised.

1088 records are indexed under the term lung diseases.

The user now combines search statements 1 and 2. He could have started by entering air pollution and lung diseases as a single search term (all Boolean combinations including negation--not--may be used). Note that he can now refer to air pollution as 1, lung diseases as 2, the combination of them as 3, etc. There are 21 records that meet the two search criteria.

The user enters a PRINT command requesting that two records be printed. This gives him three elements--AU, TI, SO--(an arbitrary selection, modified for each data base) of the first two items that qualify. He could have specified more elements, fewer elements, or all elements.

```
HELLO FROM ORBIT II.  YOUR DATA BASE WAS LAST UPDATED IN DEC 1970.
PLEASE TYPE YOUR IDENTITY CODE AND STRIKE THE CARRIAGE RETURN KEY:
USER:  ordex473
PROG:
FOR MORE INFORMATION STRIKE THE SPACE BAR AND THEN THE CARRIAGE RETURN
KEY.  OTHERWISE ENTER SEARCH STATEMENT 1, OR A COMMAND.
USER:  "explain unit record"
PROG:
THE UNIT RECORD FOR EACH DOCUMENT CONTAINS THE FOLLOWING INFORMATION:

-------------------------------------------------------
ABBREVI-     CATEGORY                  SEARCHABLE
ATION        NAMES                     CATEGORIES
  AU         AUTHORS                       X
  TI         TITLE
  AN         ACCESSION NUMBER              X
  PD         PUBLICATION DATE              X
  SH         SUBJECT HEADINGS              X
  CN         CLASSIFICATION NUMBER         X
  LA         LANGUAGE OF PUBLICATION
  SO         SOURCE JOURNAL                X
-------------------------------------------------------
SS1/C?---SEARCH STATEMENT 1 OR COMMAND--
USER:  air pollution
PROG: PSTG---NUMBER POSTINGS (423)
SS2/C?---SEARCH STATEMENT 2 OR COMMAND--
USER:  lung diseases
PROG: PSTG---NUMBER POSTINGS (1088)
SS3/C?---SEARCH STATEMENT 3 OR COMMAND--
USER:  1 and 2
PROG: PSTG---NUMBER POSTINGS (21)
SS4/C?---SEARCH STATEMENT 4 OR COMMAND--
USER:  "print 2"
PROG:

AU - LEFCOE NM
AU - WONNACOTT TH
TI - THE PREVALENCE OF CHRONIC RESPIRATORY DISEASE IN MALES.
SO - CANAD MED ASS J 102 381-5 28 FEB 70

AU - LILLINGTON GA
```

FIGURE 4 Sample ORBIT Printout

```
TI - NITROGEN DIOXIDE -- THE NEW "YELLOW PERIL".
SO - JAMA 212 1368 25 MAY 70
HOW MANY MORE ON-LINE?
USER:  none
SS4/C?---SEARCH STATEMENT 4 OR COMMAND--
USER:  brain#
PROG: MM-MULTI MEANING (BRAIN#) (3) TERMS
SEE? (YES/NO/ALL)
USER:  "?"
PROG:
SEE? (YES/NO/ALL) - DO YOU WANT TO SEE THESE TERMS?  ANSWER 'YES',
'NO' TO CANCEL THE REQUEST, OR 'ALL' TO RETRIEVE ON ALL OF THEM.
USER:  yes
PROG:
1  BRAIN (SH)
2  BRAIN STEM (SH)
3  BRAIN RV (AU)
NUMBERS, ALL, OR NONE--SPECIFY BY LIST OF NUMBERS, ALL, OR NONE.
USER:  none
SS5/C?---SEARCH STATEMENT 5 OR COMMAND--
USER:  "neighbor dure)"
PROG:
 POSTINGS        TERM
    29           DURA MATER (SH)
     1           DURAND DP (AU)
     3           DURELL J (AU)
     1           DUREY D (AU)
     4           DURLACH J (AU)
DUREL ITSELF DOES NOT EXIST IN THE INDEX
SS5/C?---SEARCH STATEMENT 5 OR COMMAND--
USER:  "tree air pollution"
PROG:
  ENVIRONMENTAL HAZARDS
    AIR POLLUTION
      AIR POLLUTION, RADIOACTIVE
        AUTOMOBILE EXHAUST
        DUST
SS5/C?---SEARCH STATEMENT 5 OR COMMAND--
USER:  pregnancy and rubella or pregnancy and measles
PROG: PSTG---NUMBER POSTINGS (231)
SS6/C---SEARCH STATEMENT 6 OR COMMAND--
USER:  "print off-line"
```

The user is asked how many more records he wants to see on-line. However, he could have specified more in the first place or could have had other than the first two printed. He could also have had records from any previous search statement printed.

The # sign is an "any character" symbol and may be used at the end or in the middle of a term to stand for any missing character(s). This is especially useful for very long or incompletely known terms.

The user is unsure of the meaning of the last message and types in ? as a command. A brief explanation is then provided.

In this instance there are three terms that qualify, two subject headings referring to anatomical parts and an author named Brain. The user can now indicate which Brain he wants to search on. If he is still unsure, he can retrieve one or more items from each and display their titles, subject headings, and other elements.

The NEIGHBOR command retrieves the alphabetic neighbors of the search term in the index. Five are automatically displayed unless the user asks for more. This command is useful for uncertain spellings and for browsing.

The TREE command is a special command used with hierarchically structured data bases. It prints the superordinate class and all the terms in the first subordinate class. By moving up or down the tree the user can quickly determine what logically related terms may be in the index.

The user instructs the program to print the 231 records off-line. This procedure is frequently used when there are many records to print, usually after a few of them have been explored.

FIGURE 4 (Continued) Sample ORBIT Printout

RECON/STIMS

1. GENERAL INFORMATION

1.1 System Name -- RECON/STIMS (Remote Console/Scientific and Technological Information Modular System) or simply RECON. A nearly identical but proprietary version is called DIALOG.

1.2 Source -- RECON software written by Lockheed Missile & Space Company, Sunnyvale, California, and STIMS software written by Informatics TISCO, Bethesda, Maryland, for NASA.

1.3 Plans for Maintenance & Improvement -- NASA will maintain and improve both RECON and STIMS at the NASA Scientific and Technical Information Facility. Improvements will center on communications (by using a front-end communication's processor), capacity (additional terminals), and new commands (numeric range search).

1.4 Type of Support -- NASA is now entering into a maintenance and computer service contract with TISCO.

1.5 Availability -- Yes, it is a government-owned system available from COSMIC,* University of Georgia, Athens, Georgia.

1.6 Cost -- Government-owned. There will be a charge of $59.00 for STIMS documentation and $14.50 for RECON documentation.

1.7 User Population -- European Space Research Organization, Atomic Energy Commission, Department of Justice, Library of Congress.

1.8 Source Language -- On-line programs are written in basic assembly language. Batch programs are written in PL/1 except for the Master I/O Control Programs which are written in basic assembly language.

1.9 Proprietary Software -- No.

---

* COSMIC (Computer Software Management and Information Center) was established early in 1966 at the University of Georgia to collect and disseminate to the public computer software developed by government agencies.

1.10 Documentation -- (1) RECON Operation Manual

(2) RECON Programming Documentation

(3) STIMS File Maintenance Subsystem

2. OPERATIONAL ENVIRONMENT

2.1 Hardware (minimum configuration)

2.1.1 Main Frame -- IBM 360/50

2.1.2 Input Devices -- Card reader or tape for batch and CRT with keyboard for on-line mode.

2.1.3 Output Devices -- 1403 high speed printer with an upper and lower case print train on the central computer and a local printer at each terminal.

2.1.4 Mass Storage Devices -- Disk and data cells.

2.1.5 Document Storage Devices -- Microfiche (manually retrieved).

2.1.6 Communication Equipment -- 25 terminals consisting of CRT, keyboard and printer.

2.1.7 Core Size -- RECON requires 150,000 bytes and STIMS requires 200,000 bytes. Another 3,000 bytes are required for each terminal being serviced.

2.2 Operating System Version -- 360/OS under MFT II.

2.2.1 Mode of use -- On-line and batch.

3. SOFTWARE CHARACTERISTICS

3.1 Operating System Environment -- IBM 360/OS under MFT II.

3.2 Transferability between Hardware -- Within IBM 360.

3.3 Transferability between Operating Systems -- Within 360/OS.

3.4 Type of Security -- No security is available except by terminal.

3.5 Back-up Facility -- Data back-up by a tape dump.

3.6 Restart & Recovery Capability -- Yes.

3.7 System Statistics -- Batch run available to get data-base statistics to find out whether the files should be reorganized.

3.8 Selective Dissemination of Information -- There are two ways to handle S.D.I. in the system. One way is by restricting the search to a range of access numbers and achieving the effect of searching only the current tape. Another way is to create a new temporary inverted file for new documents and perform S.D.I. search only on the new inverted file.

3.9 Indexing -- Manual.

3.10 Thesaurus -- There is a thesaurus used for input quality control and also for searching from remote consoles. There are five cross-references being defined: broader term, narrower term, related term, use and use for.

3.11 Input Data Editing & Validation -- Yes, there is a thesaurus file used for input quality control.

3.12 Linkage to User Code -- No.

4. USER INTERFACE

4.1 Data Description Language -- RECON/STIMS has a data definition facility. There are two sets of tables: file description table and field description tables.

4.2 Query Language -- There are two query languages: batch-mode queries and on-line queries. In the on-line system, it is possible to search only by using inverted index terms as part of the query. In the batch-mode search, one may use not only the inverted index terms but any field in the record. The following is the description of the on-line query language:

Device -- CRT, keyboard.

Language Type -- Command type with verbs followed by list of parameters.

Arithmetic Capability -- None.

Boolean Logic for Selection -- All of the logical connectors.

Selection via Ranges of Value -- Yes.

Invocation of Predefined Queries -- Queries may not be predefined and kept for the on-line (RECON) system, however, the facilities exist in the batch (STIMS) system.

Sample -- Sample system commands consist of "EXPAND", "SELECT", "COMBINE", "DISPLAY", "PRINT", "TYPE", "KEEP", "END SEARCH", "LIMIT", etc.

4.3 Output Report Language -- It is part of the query language. The output contains microfiche location codes and the microfiche documents may be retrieved manually.

Device -- Teletypewriter, display station.

Language Type -- Same as query language.

Prestored Format -- There exists a list of standard output formats. The user may modify only one of these formats for his own special use.

On-line or Off-line Print Command -- Yes.

Sort Specification -- It is not possible to sort the output of an on-line query but sorting may be specified using batched query.

4.4 Maintenance & Update Language -- There is no on-line file maintenance. However, it is possible to do updating simultaneously with searching by submitting maintenance in the background. There exist lock-out bits in a record to prohibit access to a record while updating. Language form is unknown.

4.5 Browsing Language -- Citations, abstracts or full text may be scanned on the CRT. Command language allows paging through a document or skipping to next retrieved item.

5. INTERNAL ORGANIZATION

5.1 Data Base -- There are two sets of files: the linear file which is the main file ordered by accession number, and the inverted files. NASA has 5 inverted files: descriptors, authors, cooperative authors, report numbers, and contract numbers.

5.2 Data Structure -- The record structure consists of a fixed length header followed by a variable number of variable length fields. Each field has a tag and a count associated with it. No hierarchy is permitted in the record structure.

5.3 Storage Structure -- The disk space is organized to permit variable length logical records blocked equal to track size. At the end of each block or record (if the record is bigger than one track) there is an expansion area for record overflow. There are indexes at the track and cylinder level plus an additinal master index. Records within a track are packed and maintained in sequential order.

6. OPERATIONAL FUNCTIONS

6.1 Data Access Method -- NASA has programmed its own version of a blocked, variable length ISAM (Index Sequential Access Method).

6.2 Search Strategies -- Index sequential search of inverted files.

6.3 Update Facilities -- It is possible to update in a batch mode in the background while searches are being conducted in the foreground.

6.4 Response Time

6.4.1 Search Response -- With 15 terminals running, the response is approximately 15 to 20 seconds.

6.4.2 Update Time -- It takes 0.06 seconds to change a field in an existing record.

6.5 Space -- The system imposes no maximum record size. The inverted indexes in the current file occupy about one-sixth of the space devoted to the main file. There are now 750,000 accessions in the file requiring approximately 800 bytes each.

☆ U. S. GOVERNMENT PRINTING OFFICE : 1971—481-326/81

FORM NBS-114A (1-71)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET	1. PUBLICATION OR REPORT NO. NBS TN-599	2. Gov't Accession No.	3. Recipient's Accession No.

4. TITLE AND SUBTITLE

A Survey of Selected Document Processing Systems

5. Publication Date October, 1971

6. Performing Organization Code

7. AUTHOR(S) Elizabeth Fong

8. Performing Organization

9. PERFORMING ORGANIZATION NAME AND ADDRESS

NATIONAL BUREAU OF STANDARDS
DEPARTMENT OF COMMERCE
WASHINGTON, D.C. 20234

10. Project/Task/Work Unit No. 6401137

11. Contract/Grant No.

12. Sponsoring Organization Name and Address

Same as No. 9.

13. Type of Report & Period Covered Final

14. Sponsoring Agency Code

15. SUPPLEMENTARY NOTES

16. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.)

There are many document processing systems that are commercially available or government-owned. These systems emerged in the evolution from early efforts in library automation to current on-line systems. Due to the diverse nature of the facilities provided in the document processing systems, it is difficult to evaluate them. The purpose of this paper is to present a list of features as a set of dimensions along which to compare the surveyed systems. The feature list is also developed to serve as a common basis for describing document processing systems. Another purpose of this paper is to provide a reference tool for the eight systems surveyed. They are CIRCOL, DDC ITIRC, The Mead Data Central, MEDLARS II, New York Times Information Bank, ORBIT II, and RECON/STIM. This paper first explored the characteristics of available, large document processing systems in general. An overview of the eight systems surveyed is presented. The paper then defines the feature list. The description of the eight systems surveyed according to the feature list outline is included as an Appendix.

17. KEY WORDS (Alphabetical order, separated by semicolons) Bibliographic system; computer package; data base; document processing; information retrieval; document storage and retrieval; text; text processing.

18. AVAILABILITY STATEMENT

[X] UNLIMITED.

[ ] FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NTIS.

19. SECURITY CLASS (THIS REPORT) UNCLASSIFIED

21. NO. OF PAGES 67

20. SECURITY CLASS (THIS PAGE) UNCLASSIFIED

22. Price 65¢

USCOMM-DC 66244-P71

NBS TECHNICAL PUBLICATIONS

PERIODICALS

JOURNAL OF RESEARCH reports National Bureau of Standards research and development in physics, mathematics, chemistry, and engineering. Comprehensive scientific papers give complete details of the work, including laboratory data, experimental procedures, and theoretical and mathematical analyses. Illustrated with photographs, drawings, and charts.

Published in three sections, available separately:

• Physics and Chemistry

Papers of interest primarily to scientists working in these fields. This section covers a broad range of physical and chemical research, with major emphasis on standards of physical measurement, fundamental constants, and properties of matter. Issued six times a year. Annual subscription: Domestic, $9.50; $2.25 additional for foreign mailing.

• Mathematical Sciences

Studies and compilations designed mainly for the mathematician and theoretical physicist. Topics in mathematical statistics, theory of experiment design, numerical analysis, theoretical physics and chemisty, logical design and programming of computers and computer systems. Short numerical tables. Issued quarterly. Annual subscription: Domestic, $5.00; $1.25 additional for foreign mailing.

• Engineering and Instrumentation

Reporting results of interest chiefly to the engineer and the applied scientist. This section includes many of the new developments in instrumentation resulting from the Bureau's work in physical measurement, data processing, and development of test methods. It will also cover some of the work in acoustics, applied mechanics, building research, and cryogenic engineering. Issued quarterly. Annual subscription: Domestic, $5.00; $1.25 additional for foreign mailing.

TECHNICAL NEWS BULLETIN

The best single source of information concerning the Bureau's research, developmental, cooperative, and publication activities, this monthly publication is designed for the industry-oriented individual whose daily work involves intimate contact with science and technology—*for engineers, chemists, physicists, research managers, product-development managers, and company executives.* Annual subscription: Domestic, $3.00; $1.00 additional for foreign mailing.

NONPERIODICALS

Applied Mathematics Series. Mathematical tables, manuals, and studies.

Building Science Series. Research results, test methods, and performance criteria of building materials, components, systems, and structures.

Handbooks. Recommended codes of engineering and industrial practice (including safety codes) developed in cooperation with interested industries, professional organizations, and regulatory bodies.

Special Publications. Proceedings of NBS conferences, bibliographies, annual reports, wall charts, pamphlets, etc.

Monographs. Major contributions to the technical literature on various subjects related to the Bureau's scientific and technical activities.

National Standard Reference Data Series. NSRDS provides quantitative data on the physical and chemical properties of materials, compiled from the world's literature and critically evaluated.

Product Standards. Provide requirements for sizes, types, quality, and methods for testing various industrial products. These standards are developed cooperatively with interested Government and industry groups and provide the basis for common understanding of product characteristics for both buyers and sellers. Their use is voluntary.

Technical Notes. This series consists of communications and reports (covering both other agency and NBS-sponsored work) of limited or transitory interest.

Federal Information Processing Standards Publications. This series is the official publication within the Federal Government for information on standards adopted and promulgated under the Public Law 89–306, and Bureau of the Budget Circular A–86 entitled, Standardization of Data Elements and Codes in Data Systems.

Consumer Information Series. Practical information, based on NBS research and experience, covering areas of interest to the consumer. Easily understandable language and illustrations provide useful background knowledge for shopping in today's technological marketplace.

NBS Special Publication 305, Supplement 1, Publications of the NBS, 1968-1969. When ordering, include Catalog No. C13.10:305. Price $4.50; $1.25 additional for foreign mailing.

Order NBS publications from:

Superintendent of Documents
Government Printing Office
Washington, D.C. 20402

U.S. DEPARTMENT OF COMMERCE
National Bureau of Standards
Washington, D.C. 20234

OFFICIAL BUSINESS

Penalty for Private Use, $300

POSTAGE AND FEES PAID
U.S. DEPARTMENT OF COMMERCE