FPMR 101—11.8

RECORDS MANAGEMENT HANDBOOK

Managing Information Retrieval

INFORMATION RETRIEVAL

1972

GENERAL SERVICES ADMINISTRATION
NATIONAL ARCHIVES AND RECORDS SERVICE
OFFICE OF RECORDS MANAGEMENT

Federal Stock Number

FOREWORD

Management at every level is being subjected to increasing pressure to automate the files of the office—to adopt new, nonconventional methods and equipment to improve the dissemination, storage, and retrieval of information. Professional journals, trade magazines, and agency publications are constantly reporting how management is solving its information problems through the use of these new systems. But today's manager knows that the new systems usually represent a sizable investment, and he is also aware that the investment has not always paid off.

It is the purpose of this handbook to provide the manager and those who assist him with guidelines for determining where these new systems might profitably be employed in Government offices and with criteria for selecting the right methods and equipment. While the main objective is to encourage greater use of modern information retrieval techniques, the guidelines should also help prevent the installation of ill-advised or unprofitable systems. For those offices that have already installed modern information retrieval systems, the handbook may prove helpful in analyzing and evaluating existing system performance or in revising an ineffective system.

This handbook is intended primarily for the use of management analysts, systems personnel, middle management, and any others who may be directly involved in conducting information retrieval studies or in designing and installing an information retrieval system.

Although this handbook is issued as one of a series of Records Management Handbooks produced by the National Archives and Records Service, General Services Administration (GSA), the United States Air Force shared in its development. It was produced under a contract jointly funded and administered by the Air Force and GSA.

CONTENTS

I. WHY NEW INFORMATION RETRIEVAL SYSTEMS ARE NEEDED

Conventional methods for storing and retrieving information have been doing an effective information handling job for some 50 years, and in many situations today are still the best answer. However, during and since World War II more and more people have been questioning these conventional methods and looking for new and better ways to satisfy their information needs. The three main reasons for this exploratory research have been the information explosion, the trend toward a much higher degree of specialization in all technical fields, and the advent of the new technologies of electronic data processing and document miniaturization.

The information explosion is now overtaxing conventional methods and equipment for indexing and storing the thousands of new documents being prepared each year. The trend toward greater specialization is resulting in preparation of documents that deal with increasingly narrow aspects of subject topics. New classes of information are constantly being formed by the emergence of interdisciplinary specialists. Conventional methods for classifying and indexing information are frequently not well suited to meet the demands for greater specificity in organizing and retrieving information nor the need to manipulate information freely.

Information specialists in the scientific and technical fields were among the first to apply the electronic computer, microforms, and other nonconventional methods and equipment to solve information retrieval problems. This handbook draws largely on their knowledge and experience.

What Is Information Retrieval?

It is the approach to the problem of information dissemination, storage, and retrieval that is new—nonconventional methods and equipment that have been introduced during the last decade or so. It is this new, nonconventional approach which has become known as "information retrieval." Stated in other ways:

- Information retrieval employs methods and equipment that depart in one way or another from the conventional methods we find in most offices and libraries.
- Information retrieval means there are now available methods and equipment for disseminating, storing, and retrieving information that make it possible, and often quite practical, to do things that no one considered doing before.
- Information retrieval means simply new ways for performing old tasks and is used primarily when conventional methods will no longer suffice.

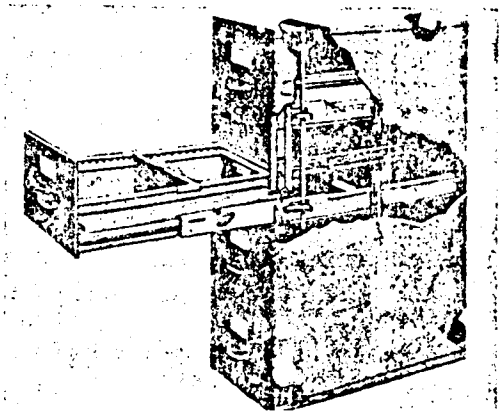
Perhaps one of the best ways to define nonconventional systems is to first explain what is meant by conventional methods and equipment—hence, the things not covered in this handbook. Examples of these conventional methods are shown in figure 1, which includes a standard file cabinet, a reference visible file, a mobile shelf file, a rotary file, and a mechanized file.

Summary of Conventional Methods

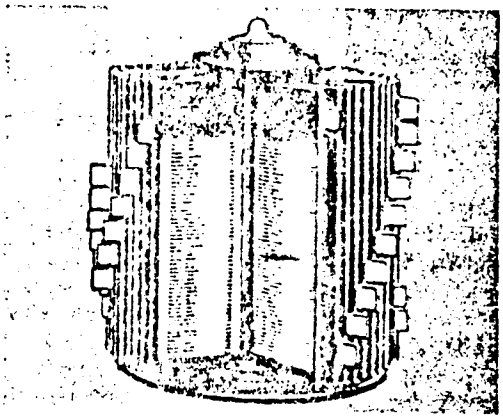
The characteristics of the documents and the methods used in organizing the information in conventional files are as follows:

- The documents are largely in paper form.
- The documents are maintained in a structured file, that is, a file organized and arranged for direct searching according to the filing feature (name, number, subject, etc.) most often known by the user when looking up the information.

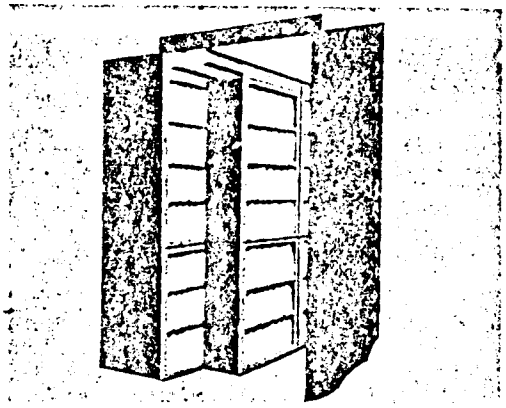
EQUIPMENT FOR CONVENTIONAL FILES



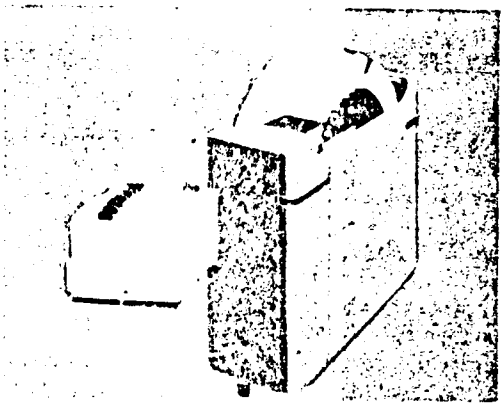
Vertical File Cabinet



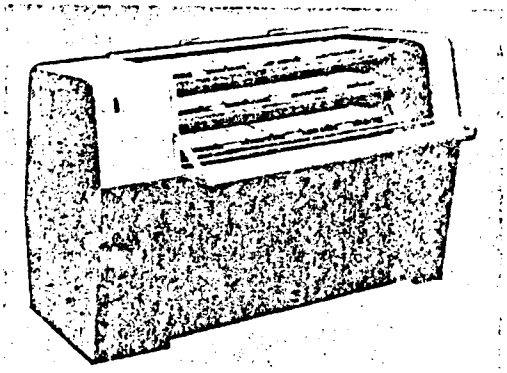
Reference Visible File



Mobile Shelf Files



Rotary File



Mechanized Horizontal File

Figure 1

- If necessary, separate manual index files or finding aids are maintained to help find information when users ask for it on a basis different from that by which the document file is structured.

The success of conventional methods depends largely on the following factors:

- Stability of information and language contained in the documents.
- Simplicity and shortness of the documents.
- Predictability of users' needs and the way in which they will ask for documents.
- Simplicity of users' needs.
- Availability of space close to the users to store the documents.

The following GSA-Records Management Handbooks relate primarily to conventional systems and should be carefully reviewed before any information retrieval study is undertaken:

Files Operations—FSN 7610-985-6973—1964

Subject Filing—FSN 7610-926-2128—1966

File Stations—FSN 7610-926-2129—1967

Summary of Nonconventional Methods

Nonconventional methods for storing and retrieving information have one or more of the following characteristics:

- The information is disseminated and stored in miniaturized form.
- The document file is largely unstructured—the documents are filed by a simple identifier such as an accession number or machine location address.
- The contents of the documents are described in detail by means of a separate, highly manipulative index file, or the entire contents are maintained in machine-readable form.

Figure 2 shows some of the nonconventional methods and equipment employed in modern information retrieval systems and these, with a numbers of others, are described in chapters III, IV, and V of this handbook.

Edge-notched cards. Edge-notched cards have been available for many years and employ a technique that is superior to conventional filing methods in numerous applications.

Optical coincidence cards. The optical coincidence of "peek-a-boo" cards is useful in special applications for organizing and retrieving information.

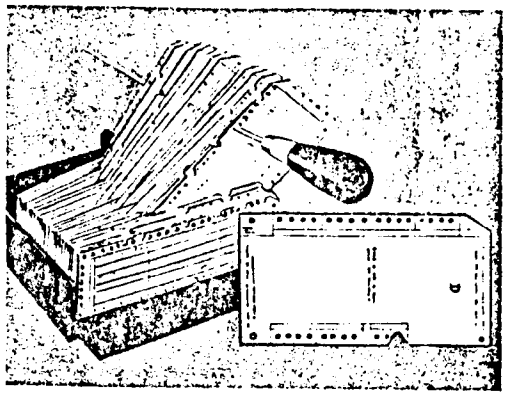
Microforms. Microfilm was conceived as a recording medium about 100 years ago, and recent developments have made microforms a vital link in solving many of today's information problems.

EAM punched cards. EAM (electrical accounting machine) punched cards have been used extensively for processing numerical data, and they can be used readily for storing and retrieving information.

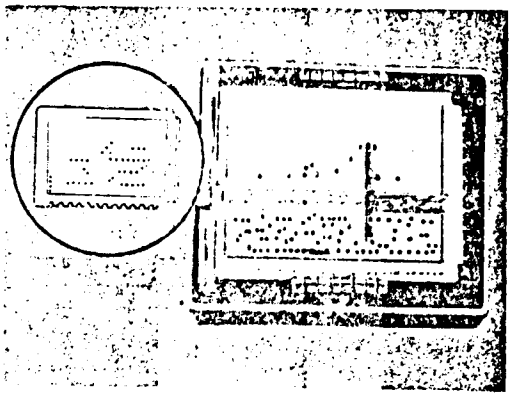
Computers. The most important of the nonconventional tools is the electronic computer, which is playing an increasingly important role in storing and retrieving information.

Nonconventional methods can often help when one or more of the following conditions exist:

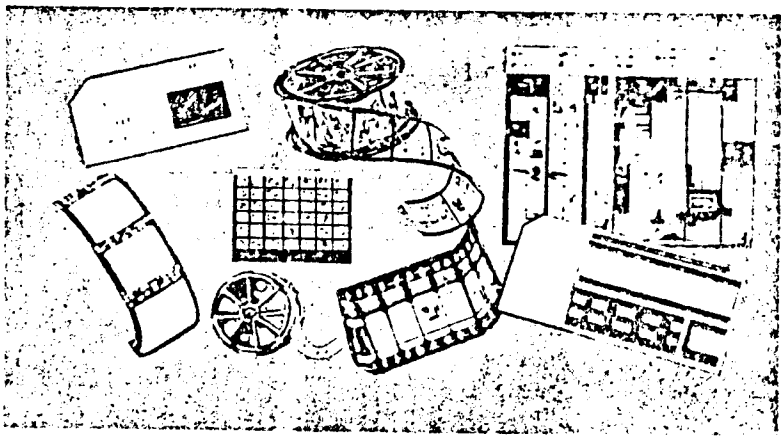
- Types of information and terminology contained in the document collection are constantly changing.
- Individual documents are lengthy and contain information on a wide variety of subjects or include large quantities of data.
- Users ask for information in a variety of ways and their needs are continuously changing.
- Users' needs are complex in that they require precise information and often must be able to correlate or manipulate it.



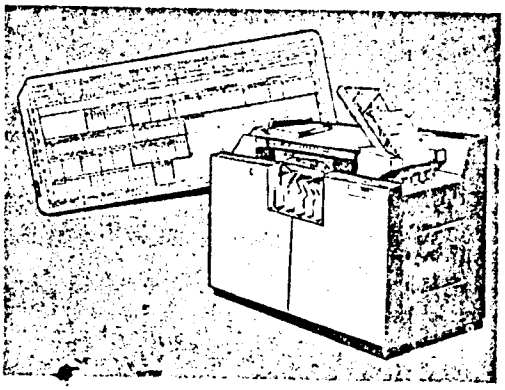
Edge-Notched Cards



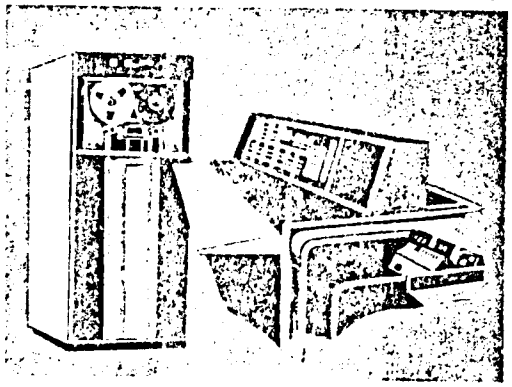
Optical Coincidence Cards



Microforms



EAM Punched Cards



Computers

Figure 2

- Documents need to be reduced in bulk or maintained in multiple sets to facilitate dissemination, storage, and retrieval.

Limitations and Advantages of Conventional Methods

To fully appreciate why nonconventional methods and equipment are needed and where they can best be used, one must first understand the sort of retrieval problems that cannot readily be solved by conventional methods. The three broad types of problems are:

- Location of specific information. Many times today the information the user needs is deeply embedded in a lengthy document—perhaps found in one paragraph of a 50-page research report. If this situation is commonplace and if there are a large number of documents in the collection, retrieval of needed information can be very difficult.

- Location of individual items of data. In some work situations it is frequently necessary to look up individual items of such data as names, numbers, dates, and amounts. The use of such conventional methods as folder files and printed listings for maintaining the data may make this a time-consuming and tiresome chore.

- Conducting coordinate-type searches. In many work situations it is necessary or desirable to conduct coordinate-type searches to identify those documents, persons, places, or things which meet a particular set of criteria. For example, management may have an urgent need for locating employees who can speak a certain language, have had certain types of experience, and are willing to travel. Conventional methods usually make it impractical, if not impossible, to conduct searches of this type.

Four general types of systems may be used for organizing information by conventional methods. The following is a description of each, together with an explanation of why each may sometimes fail.

1. Subject document files (fig. 3).

Definition: Documents arranged by subject categories, as in hierarchical subject classification

SUBJECT DOCUMENT FILE

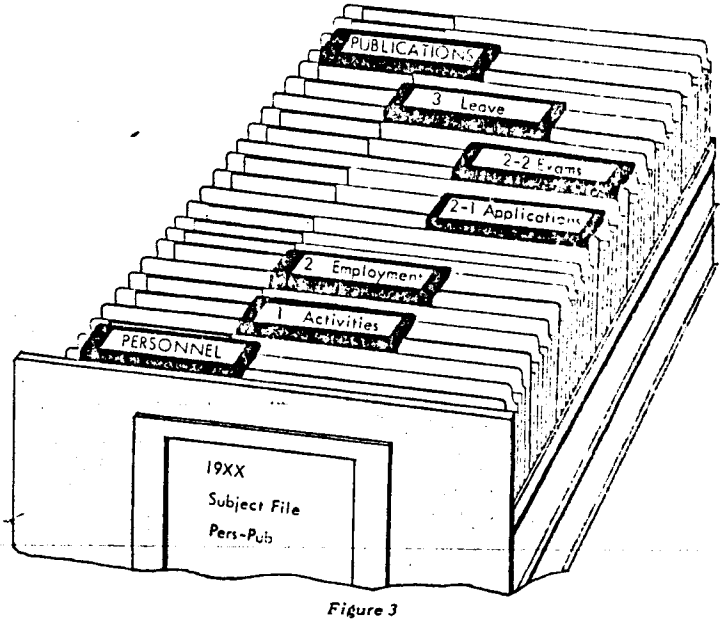


Figure 3

MANUAL SUBJECT INDEX CARD FILE

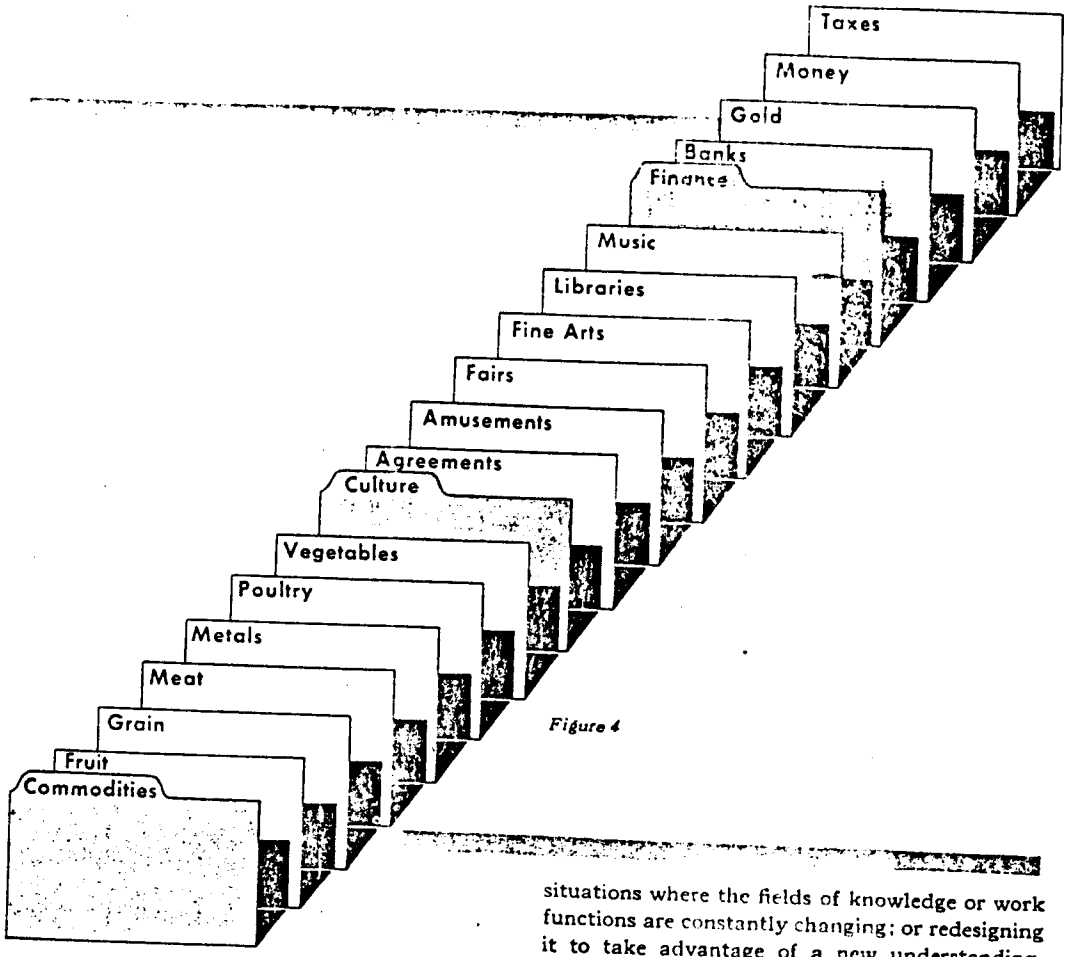


Figure 4

systems for correspondence folder files, library books, and other written material.

Significant problem: Developing a classification scheme that will satisfy the viewpoints, terminology, and needs of individual users in instances where the users have a wide variety of interests.
Why the system may fail: A hierarchical subject classification scheme needs to be directly related to the background and thinking processes of the users served. It is, therefore, virtually impossible to construct a classification scheme that will ideally serve the needs of a wide variety of interest groups.

Significant problem: Modifying the system in situations where the fields of knowledge or work functions are constantly changing; or redesigning it to take advantage of a new understanding, gained through additional experience with the system, of how the information should be organized.
Why the system may fail: Many times, the experience gained by using the system reveals shortcomings in the first arrangement that could be eliminated by reorganizing the classification structure. The rigid structure of a hierarchical classification scheme makes adjustments of this sort very difficult.

Significant problem: Classifying, filing, and retrieving individual documents in situations where they are often lengthy and involve numerous subject categories.
Why the system may fail: If an individual document relates to only one topic represented in the

subject filing system, there is no problem in filing it. But if the document has more than one subject, then cross-referencing becomes necessary. When such a situation is commonplace, the conventional system will tend to break down. A complex search involving several subjects can become a jungle of cross-references which makes the searching process very difficult, time consuming, and possibly unsuccessful.

2. Manual subject index card files (fig. 4).

Definition: Manual card files arranged by subject topics, as in a library's 3- by 5-inch subject heading card file.

Significant problem: Selecting subject terms that will be meaningful in the future.
Why the system may fail: Selecting subject topic terms that will always be meaningful and useful in the future is not only difficult but at times impossible. The problem is particularly thorny when conventional methods are employed.

Significant problem: Card preparation and updating costs.
Why the system may fail: Just the initial preparation and filing of manual index cards can be quite costly, especially if it is necessary to prepare and file several cards for each document; but to update a large file may be so costly that in actual practice it could not be done.

Significant problem: Detailed (deep) indexing of documents involving a large number of subject topics.
Why the system may fail: The physical limitations of index cards are a problem if a document must be indexed in depth. Detailed (deep) indexing of documents involving a large number of subject topics is difficult because of the size of the file that this practice would create. A card must be prepared for each subject in the document and a cross-reference prepared to all other related subjects. The structure of the card and the size of the file create barriers to fast and efficient searching. Collating these cards in a search is also very tedious and time consuming.

3. Case document files (fig. 5).

Definition: Documents arranged by case name or number, as in a personnel folder file.

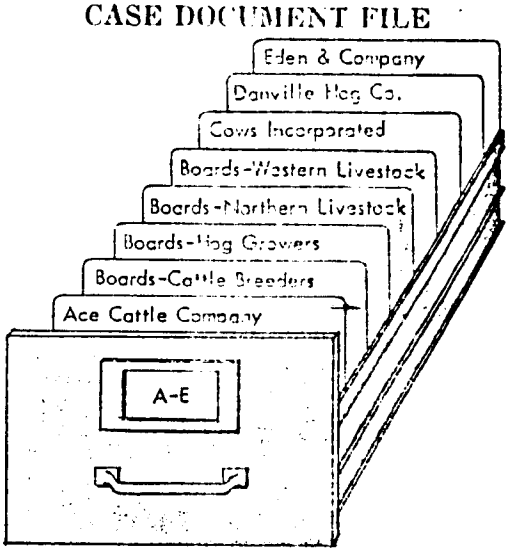


Figure 5

Significant problem: Searching large numbers of folders in situations where it is often necessary to correlate, compare, or analyze data, as in personnel selection and placement.
Why the system may fail: A case file containing large numbers of folders is very difficult to search if information must be correlated, compared, or analyzed. The physical problem of handling the folders prevents quick and easy reference. Every folder must be thoroughly analyzed from front to back before a complete job is done.

Significant problem: Locating or extracting specific items of data appearing at various places within the folder, in situations where the data is frequently needed for such purposes as answering inquiries and preparing reports.
Why the system may fail: The items of data in a document are usually not arranged for retrieval purposes but for easy preparation. When individual items must be located in a large number of case folders, the problem of pulling the folder and finding the item on the form becomes very tedious. A search of this type takes a lot of time and is subject to a large amount of human error in locating and transcribing information.

Significant problem: Locating precedent or policy material scattered among the case folders.
Why the system may fail: If material on precedent or policy matters must be located, usually it can be done only by making a search of the file

MANUAL CASE CARD FILE

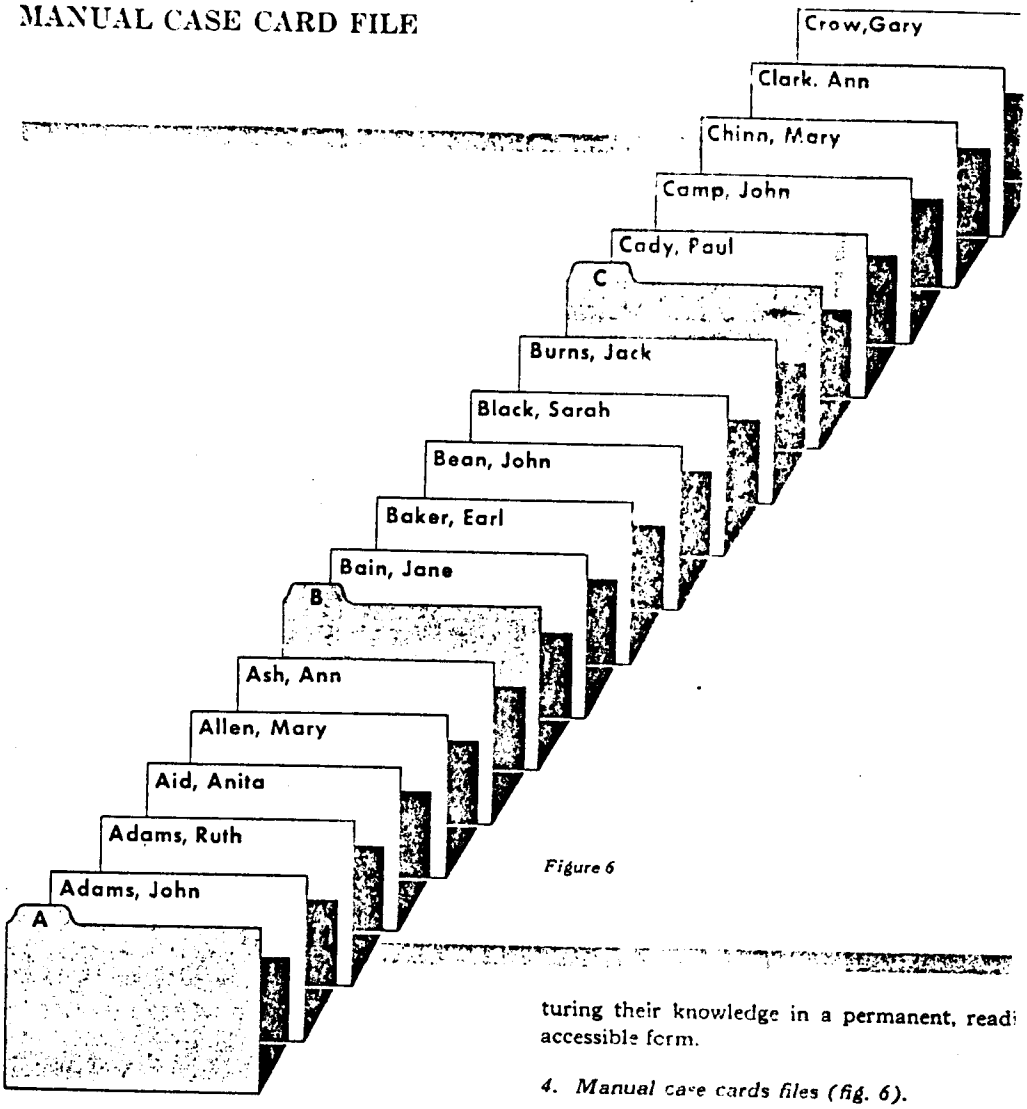


Figure 6

or calling upon the memories of employees who have had long experience in the subject matter field. Seldom is this type of information readily accessible in a separate section of the folder. The problems of interpreting precedent or policy matters are large enough; but in addition, the typical case file has the disadvantage of requiring a tiresome, page-by-page, folder-by-folder search of the file for this type of information. Many organizations that depend upon the memories of long-time employees for such information are rightfully becoming, as these older employees retire, concerned with methods and techniques for capturing their knowledge in a permanent, readi accessible form.

4. Manual case cards files (fig. 6).

Definition: Manually prepared cards arrang by case names or numbers, as in a personnel da card file.

Significant problem: Cost of updating and pr paring cards.

Why the system may fail: Card preparation a updating costs can be very high for such fil Each card must be manually prepared and ind vidually inspected. As the size of the file grow the point is reached where the cost of manual maintaining and updating the cards becom exorbitant.

Significant problem: Losing or misfiling cards.
Why the system may fail: Manual case index card files must be properly designed and controlled to prevent loss of information. If the card for a certain item is lost, then the whole record of activity for that item is lost. Although methods of color coding, grooving, tabbing, and sequential numbering can make refiling so easy that even a newcomer to the system can do the job well, most systems are not this refined. Therefore, this possibility presents a severe limitation—particularly if the information is valuable.

Significant problem: Losing vital information through illegible hand postings and errors.
Why the system may fail: Whenever a file is manually maintained, a certain loss of information results no matter how many precautions are taken to prevent it. This is particularly significant in case card files because of the uniqueness of the information placed on each card. Preparing cards in this way makes verification for accuracy a very time-consuming and costly job. The best that can be hoped for is that most of the important mistakes are found and corrected.

Conventional systems always offer certain advantages, and if they will satisfy the needs of the users, they are often preferable to nonconventional systems. Chapter VI provides guidance on how to determine which of the two methods should be used. The following are the major general advantages of conventional systems:

- Usually simpler to design and operate.
- Require no special equipment.
- Permit direct access and often facilitate browsing.
- Input costs are usually lower.

Advantages and Limitations of Nonconventional Methods

A cost-benefit study should always be made before converting from a conventional to a nonconventional system. Nonconventional methods can, under the proper circumstances and application, result in one or more of the benefits described below.

Faster retrieval. This refers to the speed at which a user gets the exact information he needs to perform a task. Fast retrieval can be the significant element of a system when need is measured in seconds or minutes. For instance, if a child has swallowed poison and the antidote must be known immediately to save a life, speed is the most essential characteristic. Or if a policeman chasing a speeding automobile calls the station to identify the license number, again fast retrieval is essential.

Better information. This means information that is more complete, more accurate, and more current. For example, modern information retrieval systems can be designed that will reduce the chance that any pertinent information will be overlooked—a most important consideration in situations such as those facing the patent attorney or physician. Modern information retrieval systems make it practical to store and correlate more information and data since they usually have the capability to reduce masses of information to a manageable proportion more quickly than conventional systems.

Conserving users' time. How much time is spent searching for information through folders, reports, card files, book indexes, and other document files in an agency or field station? No one knows exactly, but in many situations it is far too much time. In some legal offices, for example, attorneys spend as much as 75 percent of their time searching for precedent decisions and the like. Modern information retrieval methods can save valuable users' time by reducing the man-hours spent in looking up, searching for, and correlating information needed to complete their tasks. Retrieval may be simple yet time-consuming, as in looking up individual social security numbers many times each day; or again it may be as complex and time consuming as in a one-time correlation of data to determine the possible cause of a missile failure.

Improve service. This refers to providing better agency service for the general public rather than to improving service within the agency for the direct users of the information

retrieval system. In some instances, it is possible to render service never before thought possible or to improve the service far beyond that which was possible when only conventional methods were available.

The full extent of the disadvantages and limitations of a nonconventional system may not become evident until the system has been in operation for some time. This is one of the reasons that a feasibility study is needed and that careful attention must be given all aspects of the system design. (For guidance in these matters, see chapter VII.) When compared with conventional systems, nonconventional systems generally have the following disadvantages:

- Require specially trained personnel to design and operate the system.
- Usually require special equipment.
- Often require use of special procedures and techniques to retrieve information.
- Input costs are usually higher.

Coordinate Indexing—Key to Many Nonconventional Systems

The concept of coordinate indexing—or concept coordination, correlative indexing, or multiple pect indexing, as it is variously called—has be major factor in removing the restraints impo by earlier classification and indexing systems. coordinate indexing systems have one feature common: No attempt is made at time of inpu limit the description of a document by classify or indexing it under a major subject heading two. Instead, large numbers of highly definit indexing terms or data elements are employ and the document is indexed under all entries t are pertinent. To retrieve information, the u selects those indexing terms or data elements t describe the items he is looking for, and the s tem quickly identifies all those that fit his descr tion.

The key to the success of coordinate indexi is that all the descriptive information in the s tem is freely accessible, and no structuring of formation takes place until a query is receiv This permits an endless variety of on-dema searches to be made, each tailored to the preci interests and needs of the user.

Various types of equipment may be employ in coordinate indexing systems, as discussed chapters IV and V; additional information this subject is also included in chapters II and I

II. HOW COORDINATE INDEXING SYSTEMS WORK

Through the years, two traditional methods have been employed for organizing information by subject—hierarchical subject classification systems and manual subject indexing systems. The disadvantages and limitations of each were discussed in chapter I.

In hierarchical classification systems, the documents themselves are organized and arranged by primary subject categories and then further broken down by secondary categories, and so forth. Figure 7 illustrates two examples of hierarchical subject classification systems:

EXAMPLES OF HIERARCHICAL SUBJECT CLASSIFICATION SYSTEMS

Subject Numeric Filing System (office type)

ACCOUNTING
- 1 Accounts Current
- 2 Allotments
- 2–1 Symbols
- 2–2 Obligations
- 3 Disbursements
- 3–1 Loans

AUDIT
- 1 Assignments
- 2 Contract Audits

Dewey Decimal Classification System (library or office type)

600 APPLIED SCIENCE
- 610 Engineering
- 611 General Engineering
- 611.1 Equipment and Supplies
- 611.11 Tools
- 611.111 Cutting Tools
- 611.111.1 Stroke
- 611.111.11 Depth of Cut

700 ARTS AND RECREATION

Figure 7

The manual subject index file—such as the 3- by 5-inch card file found in most libraries—is often employed as a supplementary finding aid. Broad subject headings that are complete unto themselves are normally used, and the headings are arranged in alphabetical sequence. Typically, the card includes the title, date, author, and similar identifying information, perhaps plus a very brief description of the document. If the document is a book, usually several subject heading cards are prepared and filed alphabetically. Author and title cards may also be prepared and interfiled among the subject cards. The following are some examples of possible subject headings:

Automatic data processing
Correspondence management
Forms management
Information retrieval
Records retirement
Source data automation
Survey techniques
Work measurement

Principles of Coordinate Indexing

Coordinate indexing systems can be used to replace either or both of the hierarchical subject classification systems described above. The documents are identified and arranged by number, name, author, storage location address, or some other simple identifier. The index is usually a separate, highly manipulative, often mechanized file.

In a typical coordinate index, large numbers (sometimes thousands) of short terms are employed, most of which are not intended to be used alone but rather in any desired combination—"coordinated" to describe the various topics, concepts, aspects, characteristics, features, or attributes of the document or other item being indexed. These terms range from precise words and quantitative or qualitative data to abstract concepts or ideas. Both broad and narrow terms are used in the same system.

A zoologist, for example, might include such terms as those illustrated in figure 8 in his vocabulary of indexing terms:

SAMPLE VOCABULARY OF INDEXING TERMS

Africa
albatross
Antarctic
Arctic
Asia

bear
black
blue

capture
cat
color
conservation

deer
defense
diseases
dog
domestic
dorsum
duck/goose

eagle
ear
egg
elephant
Europe
exercise
exterior
eye

fish
food
fright

gestation
growth

habit
horse
housebreaking
hunting

instinct

leg
life span
lion

migration

1900 AD—present
1500–1900 AD
1000–1500 AD
Before 1000 AD

North America

obedience
offense

perception
population
preserve
price

reproduction
rescue
research
respiratory
rodent

shelter
size
South America
speed
strength

temperature—over 100°
temperature—80°–100°
temperature—60°–80°
temperature—32°–60°
temperature—under 32°

whale
white
worm

zebra
zooid

Figure 8

When indexing an individual document, all those terms that are pertinent are used to describe it. Thus, it can be seen that the description of the document consists of a group of interdependent terms that together comprise, in effect, a very brief abstract of the document.

In searching a coordinate index, one selects those indexing terms in the vocabulary that best describe the desired information. The index file is then searched to find any documents indexed under those terms.

Figure 9 illustrates the principles involved in searching a coordinate index. The cards represent indexing terms considered pertinent to a particular search question; the numbers on each card represent those documents indexed under that term. To find if there are any documents that would satisfy the search question, the searcher would look for particular document numbers that have been entered on all pertinent cards.

As in the indexing process, the searching process permits free coordination of a large number and wide variety of terms. For example, when desirable one can narrow the search by using more specific terms, or broaden the search base by dropping the more specific terms, or form new combinations of information or data by changing the configuration of the terms used in the search.

Types of Indexing Terms

Two types of indexing terms that may be used are as follows:

SEARCHING A COORDINATE INDEX

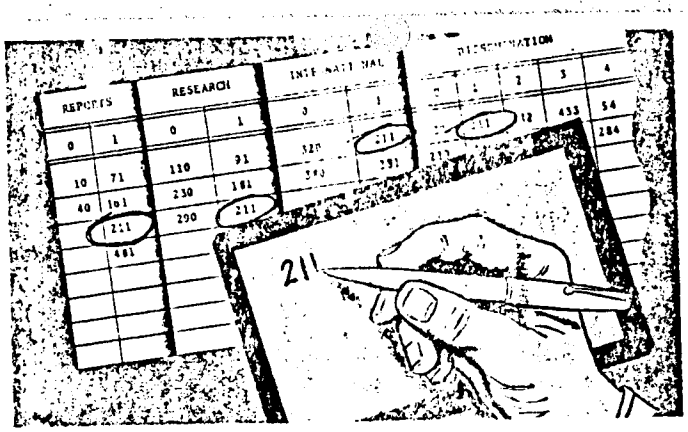


Figure 9

Keyword. The index terms consist of key words selected from the title or text of the documents. The indexing vocabulary is a by-product of the indexing process, and some form of control is usually exercised to keep the system manageable. The indexing of individual documents may be accomplished either by manual or machine (automatic) indexing methods.

Descriptor. A specially prepared vocabulary of indexing terms developed through a continuing process of analysis of the documents being indexed. The descriptors are usually formalized and controlled by means of a thesaurus. Indexing terms are manually assigned to individual documents from the approved list. Some of the terms selected to describe a particular document may coincide with keywords appearing within the document, while many will not.

Index File Arrangements

The index file is arranged in either of the two following ways:

By Document Numbers. A card or machine record is prepared for each document stored in the system, with all indexing terms describing the document recorded thereon. This is usually in coded form. Retrieval of information from the file involves sequential or serial searching, since the searcher must examine all the index records in the system to identify those documents that are assigned the terms used in the search.

By Indexing Terms. A card or machine record is established for each indexing term. When the indexer has decided which terms will be assigned to a particular document, the index records for those terms are selected and the document number is recorded thereon. Retrieval involves selective or parallel searching, since the searcher or the machine selects and examines only those records representing the terms used in the search.

Major Advantages of Coordinate Indexing Systems

- More Specific. Coordinate indexing makes it not only possible but practical to describe documents or other items in greater detail (depth) than conventional methods.

- More Adaptable. Coordinate indexes are far more adaptable to changing situations and unanticipated events than conventional methods.

- More Manipulative. Coordinate indexing makes it possible to quickly correlate and manipulate information and data in an endless variety of ways to achieve the desired search results.

Those desiring to install a coordinate indexing system have a wide variety of equipment choices. These include such manual types as the columnar, optical coincidence, and edge-notched card systems covered in chapter IV. Also, certain types of microform equipment, electrical accounting machine punched card systems, and electronic computers, described in chapters V and VI, may be used. For information about designing a coordinate indexing system, see chapter IX.

III. MICROFORM SYSTEMS

Microform is the general name for the various types and formats of microfilm and other media used for recording information in miniaturized form. In the past microform was used mainly for space-saving purposes; but numerous studies have shown that it is often less costly to place the records in the low-cost storage facilities provided by the Federal records centers. Today, however, microforms are assuming a new and far more important role in solving problems relating to information dissemination, storage, and retrieval.

How Microforms Help Solve Typical Information Problems

The following are some typical problems that can sometimes be solved or partly solved by the use of a microform. Moreover, it is not likely in any given situation that only one of these problems prevails, which largely explains the growing interest in microforms.

Problem: Document Accessibility

- Travel problem.
- Competition problem.

It is usually possible to keep near the users small collections of documents that occupy a file cabinet or bookcase. But the larger document collections, by necessity, are usually located at some distance from the users' area. This means that either the document or the user has to travel back and forth to the storage site.

Further, there are times when the same document is needed by more than one user, and each must wait his turn. These problems of course cause work delays. They also tend to reduce the usefulness of the information contained in the documents, since the users are inclined to try to do without unavailable documents if they can.

Both problems could be solved through the use of a microform system. Once the documents are converted to a microform, inexpensive duplicate sets could be placed in various locations i the users' work areas. A second choice, whic solves the competition problem only, is to mak film-to-film copies for multiple users who need t see the documents.

Problem: Document Servicing and Control

- Man-hour requirements for pulling folder and preparing document chargeouts.
- Man-hour requirements for filing returne documents.
- Man-hour requirements for following u on unreturned documents.
- Man-hour requirements for routine docu ment maintenance.

If a microform system is used, inexpensive diaz copies of the documents can be made and give to the user instead of loaning the file copy. Th user disposes of the duplicate copy when he i through with it. Thus there is no documen chargeout and refile problem, and file mainte nance is reduced to a minimum.

Because personnel costs are rising constantly and it is sometimes difficult to obtain file clerks situations will be increasing where records man agers must turn to microform to solve thei problems.

Problem: Retrieval Speed and Costs

- Random lookup of individual items o data.
- Examination of graphic information.
- Scanning and retrieving information i textual documents and indexes.

In situations where a large volume of data can be readily converted to a microform, retrieval speeds sometimes can be increased for a very

small additional cost by use of this medium. This is particularly true of instances where retrieval involves random lookup of individual items of discrete data such as a social security number, date of birth, or street address.

If there is a continuing need for examination of graphic information—such as large maps, engineering drawings, or photographs—microform often will make the job faster as well as easier. Similarly, scanning or browsing through large collections of textual material and indexes is sometimes easier and faster if they are available in microform.

Overall retrieval speeds and costs often can be improved because a microform system makes it possible to store needed documents and data at the user's work station, rather than keeping them at a remote location.

Problem: Document Printing, Distribution, and Stocking

- High costs for printing, collating, and packaging of paper documents.
- Transportation and handling costs.
- Stock control and replenishment costs.
- Time-delay problem.

Many Government agencies discovered some years ago that the most economical and efficient way to reproduce, distribute, and fill individual requests for unpublished reports is by means of the microform. Federal agencies, within the Department of Defense in particular, are saving thousands of dollars each year by using the microform for reproduction and distribution of engineering drawings of military equipment.

Not only is it sometimes possible to reduce the initial printing costs, but significant savings can often be realized in handling and transporting documents. Stocking usually can be eliminated altogether, since the microform stored at the original source or at any distribution point can be used to reproduce on demand low-cost, film-to-film copies or enlarged paper copies. The original microform can be produced readily by photographing paper documents. However, with the advent of computer-assisted document preparation, editing, index preparation, formatting, and Computer-Output Microfilm (COM) equipment, direct publication of documents in microform is now possible. The computer output magnetic tape also can be used to automatically print paper copies. For many agencies, these new techniques offer the means for a substantial reduction in the time lag between document drafting and receipt by the users.

Problem: Computer Data Storage and Accessibility

- Storage and retrieval of machine language backup data.
- Storage and retrieval of static or semistatic data.
- Direct access to computer master file.

It doesn't take long for a computer to fill a reel of magnetic tape with data. If it is kept busy all day, the computer may have produced dozens of tape reels to add to the tape library. It is little wonder, then, that some computer installations have thousands of tape reels or millions of punched cards in their file and must often restrict the computer master files to summary data. While this backup data resulting from input processing and other machine runs is usually essential to system documentation, due to its great volume it is often too costly to retain the data in machine language and search it by computer. The Social Security Administration was among the first to use the microform and the first to procure a COM device to solve this problem.

While the computer provides the fastest and most accurate means for compiling, updating, and organizing static and semistatic data, the size and cost limitations of mass memories and time requirements often make it impractical to use the computer to retrieve data from these files. Often, the best current solution to the problem is to convert data recorded on magnetic tape to a microform by means of COM equipment. A special optical mark reader, called the "Foto" Optical Sensing Device for Input to Computer (FOSDIC), has been developed to read and process Hollerith-coded data on a microfilm copy of punched cards.

Such computer data bases as inventories, transportation schedules, rates, and special tables can be converted periodically to microfilm and then searched by means of standard microfilm readers. Where static information ties in with dynamic data maintained "on-line" with the computer, special remote terminals have been designed to permit the users to interrogate both data bases at the same time.

By necessity most large Automatic Data Processing (ADP) systems must use batch-processing techniques and access the master file on a cyclical basis—perhaps once or twice a day, once a week, or possibly less frequently. During the interim, the data is locked up in the tape reels and inquiries must wait until the next processing cycle comes around to be answered. By converting the data to a microform by means of COM equipment, inquiries and requests can be handled quickly and efficiently by nonskilled personnel equipped with microfilm readers.

Problem: Updating and Maintenance of Directives, Manuals, and Catalogs

- Total costs for individual updating of directives, manuals, and catalogs kept at numerous locations.
- Errors and delays in individual updating.
- Maintaining large, frequently used manuals and catalogs intact and in good condition.

The updating of maintenance and procedural manuals, catalogs, and similar publications can be a time-consuming and difficult problem if there are numerous publications and if they are maintained at numerous locations. Errors are made in entering the changes, while the insertion of some changes is delayed or never made at all. If the manuals and catalogs receive heavy use, as they often do in a maintenance shop, the pages are likely to be torn and lost. When detailed information is needed at the job site, the mechanic may have to copy the information by hand or remove a page.

In most agencies, no one knows exactly what this is costing or is aware of the full effects of not having current, accurate data on hand at each user location. However in those instances where a detailed study was made, such as at some of the airlines, the savings were sufficient to pay for the cost of the microform system in a comparatively short time.

One of the ways to solve these problems through microform is to maintain a single master copy in cut-sheet form at a central point. Changes are entered in this master copy as they occur. The entire master copy is periodically rephotographed, reproduced in microform, and distributed to the users; whereupon, they simply dispose of the entire old copy. The microform readers are often equipped with a paper copier so that mechanics can make disposable copies to take back to their job sites when needed. In some situations the microform might also be produced through the use of the computer and COM equipment, as described earlier.

Problem: Procedural Bottlenecks

- Collection and transportation of large volumes of data.
- Verification of data on documents passing through the system.
- Logging documents.

Collection and transportation of large volumes of data such as questionnaires and reports can be a knotty problem if they are retained in their original paper form.

The U.S. Census Bureau, Department of Commerce, solves this problem by having the census questionnaires microfilmed at various locations in the field. The microfilm is then shipped to the headquarters office at Suitland, Md., where it is placed upon a FOSDIC microfilm optical mark reader. It converts the data to machine language code for processing by computers.

Several Government agencies receive large volumes of checks from the public. The checks can be microfilmed while being processed through the system in order to verify any data that may later be questioned. For similar reasons, organizations using Optical Character Recognition (OCR) equipment for computer input sometimes microfilm incoming documents.

The Division of Disbursement, Treasury Department, must maintain a record of each of the 1.5 million checks it issues each day. In the past, this was done by preparation of a paper record. Using COM equipment, the record is now produced directly from magnetic tape, making it possible to place the issue record for 102,000 or more checks on a single roll of microfilm. Duplicate microfilm copies of each month's veterans' benefit check issues are sent to Veterans Administration regional offices throughout the United States where the microfilm is used to answer thousands of inquiries a month, conduct postaudit operations, obtain a historical record of payments in specific cases, and locate addresses.

If it is necessary to log incoming and outgoing documents, microfilming is usually a much simpler and cheaper method than keeping records by hand. Many libraries use this technique for charging out books. Equipment manufacturers have developed lightweight portable cameras, including some that are battery operated, that add to the practicability of using a microform.

Problem: Storage and Handling of Large and Nonstandard-Size Documents

- Special equipment needs.
- Folding and unfolding of oversize documents.
- Storage of documents with irregular sizes and shapes.

Oversize documents such as tracings, drawings, and maps can be recorded on microfilm to eliminate the problems of special equipment requirements and the need for unfolding and folding the documents each time they are used. However, the original documents must conform to certain quality standards in order to produce a satisfactory microfilm substitute.

Documents having irregular sizes and shapes can be reduced to a uniform size through microfilm. Improved color microfilm is available if color is a significant factor.

Problem: File Integrity

- Errors in filing.
- Accidental loss or destruction of information.
- Alteration and obliteration of information.
- Users' failure to return documents.

Errors in filing occur in spite of the best efforts of file supervisors. If the file is a large one, it may be days, months, or years before a missing document turns up. Whenever a document is removed from a file and forwarded to a user, it might be lost in transit, accidentally destroyed, damaged, or not returned. These, of course, are serious risks when dealing with important documents such as those affecting individual rights and claims.

Often the best way to insure absolute file integrity is to convert documents to a microform system. The user is provided access by furnishing a film-to-film copy or an enlarged paper copy for his use.

Problem: Document Acquisition

- Rising cost of hard copy publications.
- Acquisition of rare or unique documents.

The rising costs of publications printed in paper copy are making it necessary for many libraries, offices, and others to curb their document-acquisition programs. In those instances where a document is available in either paper copy or microform, savings of 70 percent or more can usually be realized by purchasing microform.

There are also times when desired documents are out of print. If such documents are needed urgently, the simplest and generally cheapest way is to make microform copies.

Problem: Document Preservation and Protection

- Prevention of wear and defacement of valuable, irreplaceable documents.
- Protection of indispensable operating records against a disaster.
- Protection of classified documents.

The Library of Congress and the National Archives use microfilm extensively for preservation of important documents. The microfilm copies are made available to scholars and researchers, not the original documents.

Microfilm is used by many agencies for protection of indispensable operating records against a fire or national disaster. The film is usually kept in a remote, protected depository that in most instances is equipped with machines and supplies for making film-to-film copies or paper enlargements. The original copies of classified documents may be microfilmed so that either the original or copy of the document is always secure.

Problem: Equipment and Space for Document Storage

- Availability of adequate space to house documents.
- Space costs.
- Equipment costs.

While space and equipment savings are often an important factor in a microform cost-benefit analysis, microfilming can seldom be justified for this purpose alone.

Prerequisites for a Successful Microform System

For a microform to serve as a satisfactory substitute for paper copy, it must be as legible and easy to use as its paper counterpart. Microform system success depends upon such factors as condition of the original documents, the film, the camera, the camera operator's work, the quality of film processing, the suitability of the microform type, proper storage and handling of the microform, the adequacy of viewing equipment, and the ability to quickly locate information within the microform record. A weakness in any of these areas may cause the system to fail.

The single most critical factor is the condition of the document. Not only does this largely govern the quality of the finished microform, but it is a major cost factor in the filming operation. Typical problems are poor contrast between the reading matter and the paper; extremely fine lines or printing; lack of uniformity in color, size, and thickness of documents; intermingling of one-sided and two-sided documents; the need for removal of staples, pins, and other fasteners; and the need for sequence checking and screening to remove extraneous material.

Within the next 10 years it can be expected that many of the existing large-folder file systems in the Federal Government will be converted to microform. Steps should be taken as soon as possible, therefore, to clean up and revise such systems so that the essential papers will be susceptible to low-cost, high-quality microfilming. Careful attention should also be given to the planning and maintaining of any new, long-term paper document files so that they too may be readily converted to a microform should this later become desirable.

Types of Microfilm and Cameras

Normally, the initial step in any microform system is the recording of document images on roll microfilm having a silver base. This master film, in which images appear in a negative mode, is then used to produce duplicate reference copies as needed. The copies may also be silver films, but if widespread duplication is necessary the lower cost ammonia-developed diazo films are commonly used. A third type, thermally developed vesicular films, may also be used for producing reference copies.

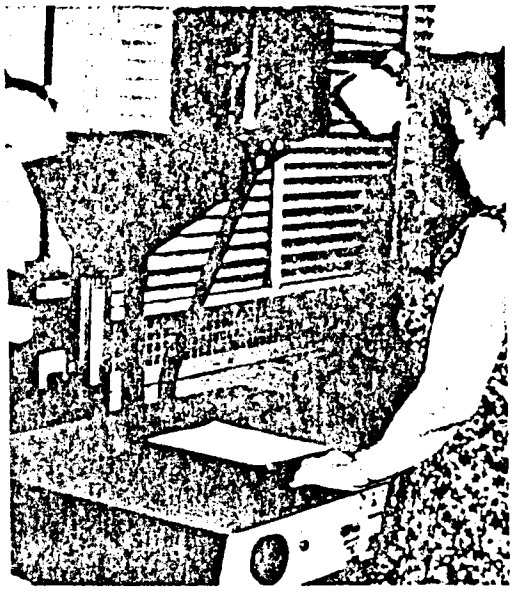
While the original microfilm master is normally in roll form of 16 mm, 35 mm, 70 mm, or 105 mm width, the reference copies are often cut into small pieces for use in systems employing unitized microform media. These include strips, chips, microfiche, microfilm jackets, and aperture cards, which are described later in this chapter.

Four main types of cameras are used in the original filming operation. See figure 10. These are as follows:

Planetary cameras are employed for obtaining high quality microfilm of engineering drawings, maps, and assorted other documents that cannot be satisfactorily filmed by a rotary camera.

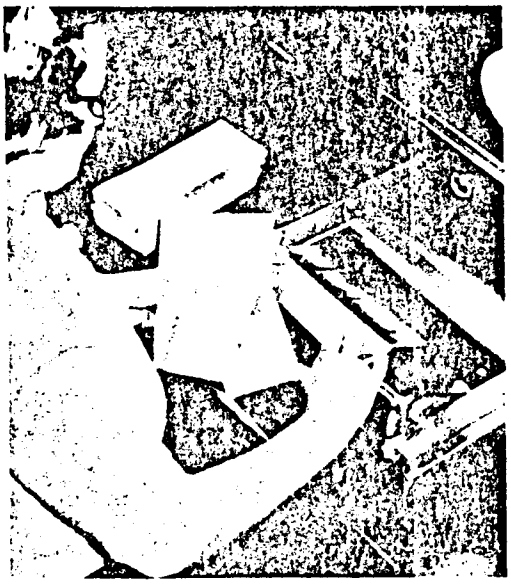
Step-and-repeat cameras are used for direct film-

TYPES OF MICROFILM CAMERAS

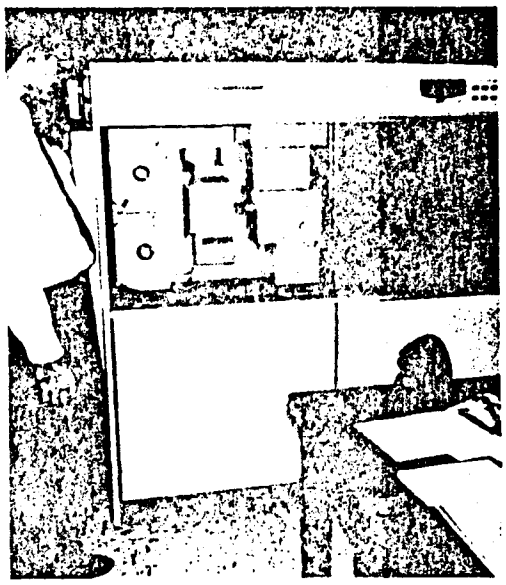


Planetary Camera

Step-and-Repeat Camera



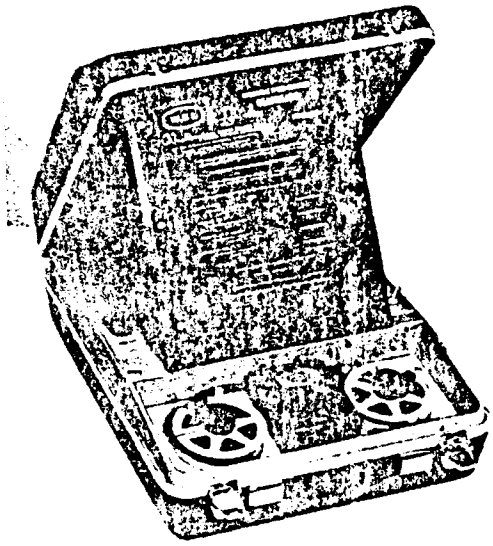
Rotary Camera



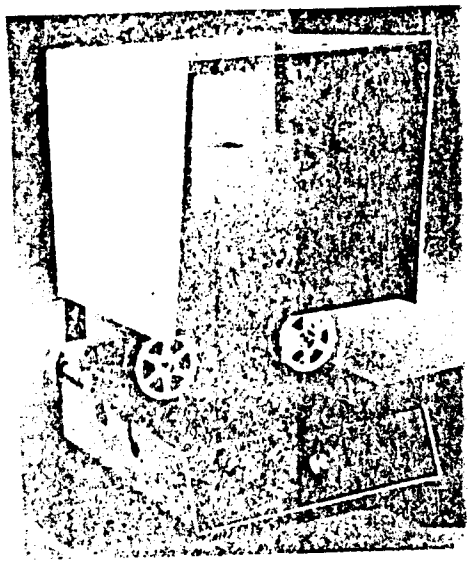
COM (Computer-Output Microfilm Device)

Figure 10

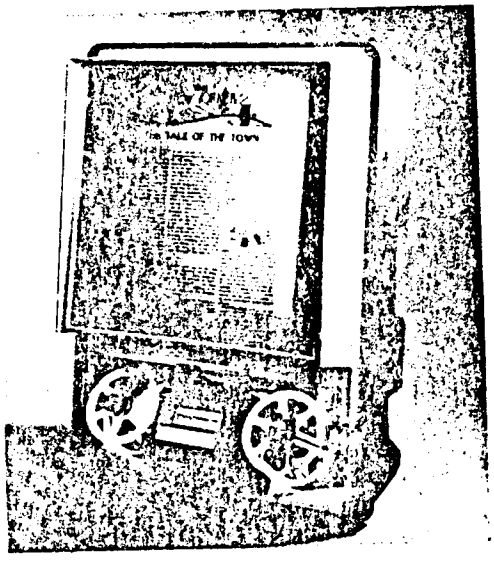
HAND-DRIVEN ROLL MICROFILM READERS



Washington Scientific Industries
Model RH Portable Reader



University Microfilms Header Model 1414

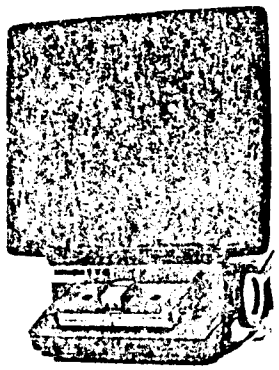


The University Microfilms
Model 1212 Reader

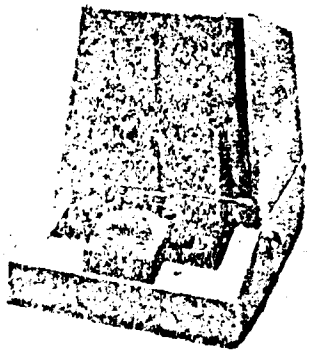
DASA Corporation's Mark I Model U Read

Figure 11

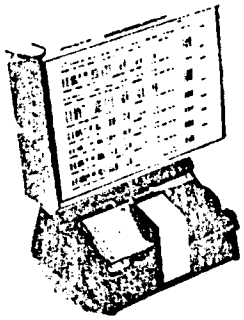
MOTORIZED READERS FOR ROLL MICROFILM



The Information Handling Services Satellite IIW Reader



The Recordak Motormatic Reader, Model MPG



The Ednalite 1624 COM Reader

The Dietzgen 4317 Reader

Figure 12

ing of documents in the multiple-row microfiche grid format. (Microfiche may also be constructed by cutting 16 mm or 35 mm film into strips and placing the strips in microfilm jackets or arranging them in rows on a special frame or sheet of clear film.)

Rotary cameras are used for filming printed and other documents of uniform size and color where ordinary film quality will suffice. They are largely automatic, thus permitting higher input speeds and use of unskilled operators.

Computer-Output Microfilm (COM) devices record computer-produced data directly onto microfilm, thereby bypassing the preparation of paper documents altogether. These devices can also add automatically to the microfilm copy the bars or code lines, image marks (blips), or photo-optical binary codes often employed to assist in the retrieval of documents or data.

Factors Affecting the Choice of the Type of Microform System

The choice of which microform system to select is governed by many factors. Mainly, these are the height and width of the documents, the number of pages per document, the total volume of documents or data, organization of the file, nature and extent of changes and additions to the file, number and location of the users, nature of

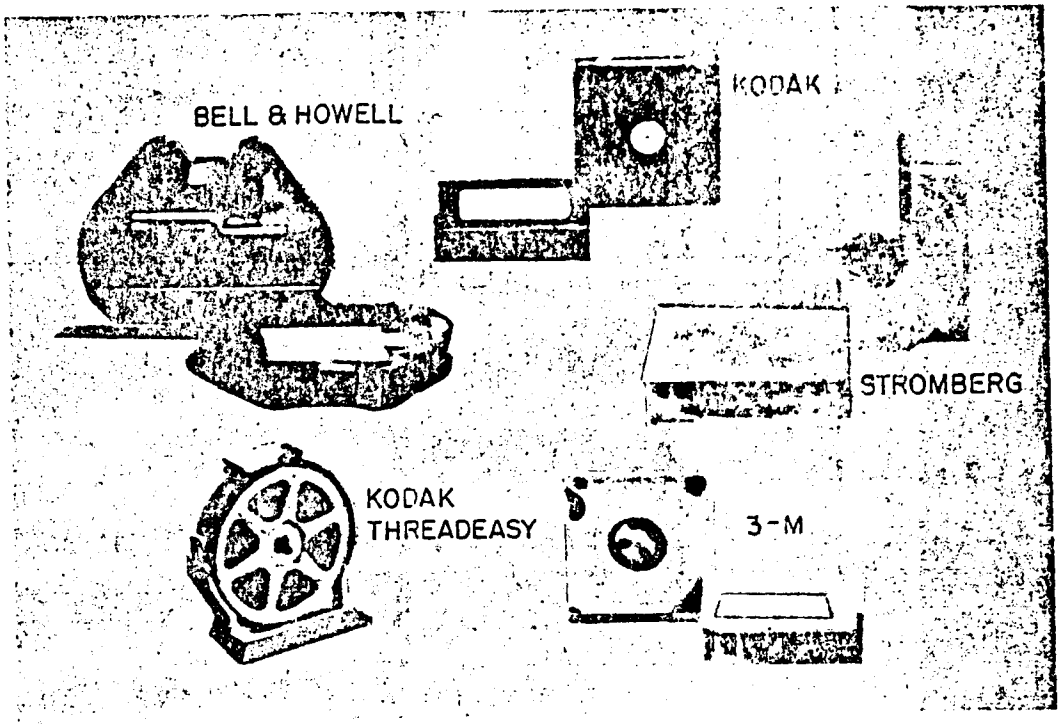


Figure 13

the reference activity, reference rate, retrieval speed requirements, and requirements for producing film or enlarged paper copies. Information on gathering the necessary data for system requirements, analyzing user needs, and selecting the right method and equipment class is provided in chapter VII.

Guidance on selection of particular manufacturer's equipment is contained in the records management handbook, *Microform Retrieval Equipment Guide*. A description of a number of systems employing microforms is included in the records management handbook, *Information Retrieval Systems*.

Types of Microform Systems

The following are descriptions of the various types of microform systems, together with a brief summary of the main advantages and limitations of each. Most of the microform readers mentioned are also available in reader-printer models that can produce full-size paper copies of the documents.

Conventional Roll Microfilm. These include systems using hand-driven microfilm readers and standard microfilm reels, as illustrated in figure 11. Flashcards or flash targets are used to separate file segments or pages. (Figure 14 depicts a sample of a flashcard used on roll microfilm.) Conventional roll microfilm systems are well suited to storage or protection of documents for archival, administrative, legal, or security purposes, and other situations where there is a very low reference activity. The main limitations of conventional roll microfilm systems are slow retrieval speeds and inconvenience to the user. The microfilm must be hand threaded through the reader, a slow and tedious operation. The user must then hand crank the film and scan the reader screen image by image until he finds the desired document.

Motorized Roll Microfilm with Mechanized Image Locator Aids. In both this and the system that follows, most of the microfilm reading equipment has been improved in three ways. First, a motor usually with both high and low speeds has been added; second, film cartridges or cassettes have been substituted for standard microfilm reels, and the reader has been made self-threading; and third, new techniques or devices have been employed to aid in locating desired film images.

Except for conventional roll microfilm systems, the motorized roll microfilm systems with mechanized image locator aids are generally the lowest in overall costs. They offer particular advantages for lengthy documents or record series. They can be successfully employed for the reproduction, dissemination, storage, and retrieval of catalogs, manuals, and publications, in which event many of the advantages described below for microfiche apply.

Figure 12 shows some typical motorized (mechanized) roll microfilm readers and reader printers while figure 13 provides examples of the various types of cartridges or cassettes employed. The mechanized image locator aids are of three types, as follows:

- Bars or code lines superimposed between images on the film that, when matched with a corresponding scale on the reader screen, can usually localize the search to within ten images or less, in a sequentially arranged numerical or alphabetical file.

- Film pull-down (linear location) aids that employ microfilm readers incorporating an odometer-like device for finding images on the basis of their linear location on the film. As in the system using image counting, this one depends upon the user's knowing or separately looking up the location of the desired image.

- Image count aids, which consist of marks (blips) superimposed beside each film image for use on a reader that has a photoelectric counting device. To locate an image, the user must know or separately look up the image location number for the document he desires to view. He enters this number on the reader keyboard, and the film automatically moves through the reader and stops when it reaches that number.

Figure 14 depicts examples of roll microfilm employing these three mechanized locator aids.

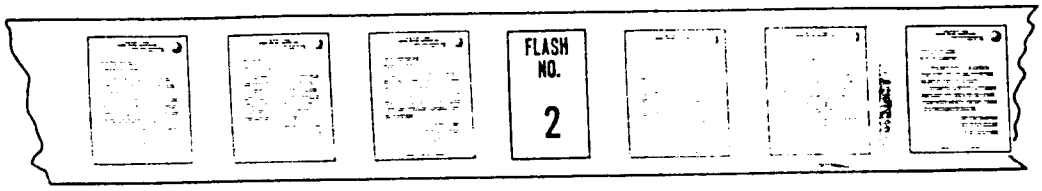
The use of the cartridges and cassettes with self-threading motorized microfilm readers has substantially improved the ease and convenience in the use of roll microfilm. The image-finding aids are a real boon to retrieval speeds in situations where they can be satisfactorily applied. Of the three techniques, the film pull-down (linear location) is usually the least costly and can be incorporated into a system quite easily. The bar or code systems are the next least costly and somewhat more difficult to incorporate into a system.

All three image-finding techniques have certain limitations. Bar or code line systems can be used only where the file is sequentially arranged by numerical or alphabetical identifiers and the user is conducting his search on the same basis. While the film pull-down (linear location) and image count techniques permit the documents to be in random sequence, a separately maintained list or index may be required for use in determining the proper microfilm roll and image location. Systems employing the image count technique require microfilm readers that are more complex and hence normally more costly than those used in the other two.

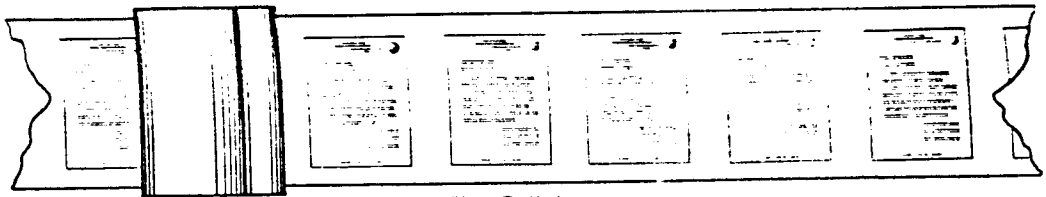
Special Note on Changing or Adding to Roll Microfilm

Most roll microfilm systems have one problem in common—changing or adding to previously filmed records. There are three methods for doing this, and none may prove entirely satisfactory. However, under certain circumstances, one or more might prove practical. The first and least likely method (except for publication of catalogs, manuals, listings, and COM produced items) is to retain the original documents, make the changes, and periodically refilm the entire file. A second but not always practical choice is to film the changes or additions and splice the new film onto the old film. A third method is to film the changes

INDEX METHODS USED IN 16 mm FILM



Flash Card



Film Pull-down

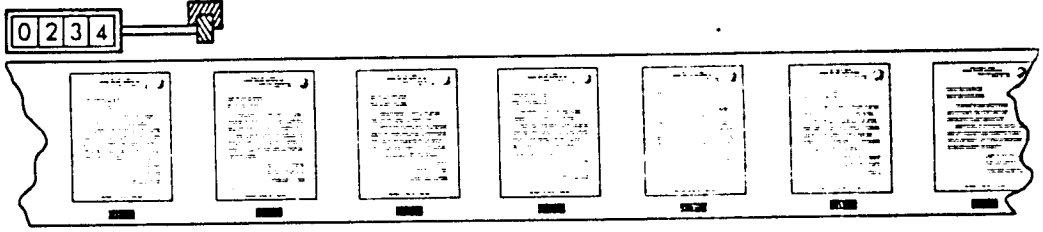
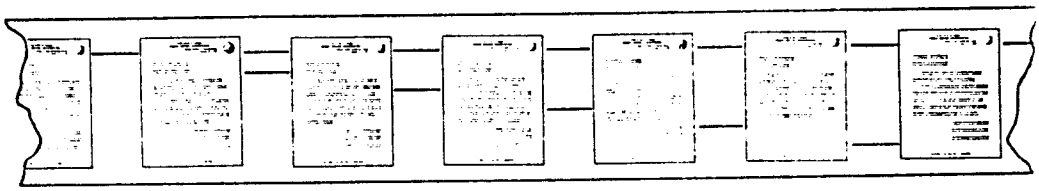


Image Count



Bar or Code Line

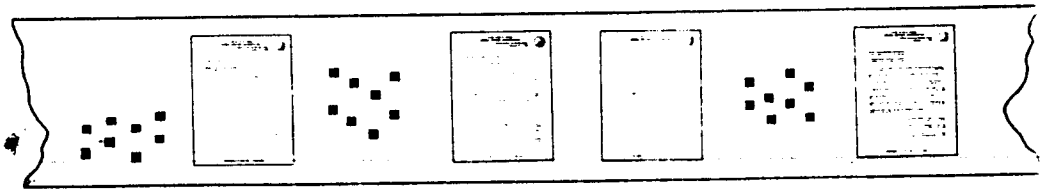


Photo-optical Code

Figure 14

and additions as they are received, add the new rolls to the microfilm collection, and maintain a separate index or locator record (preferably computer maintained and produced) showing the location (microfilm roll and possibly image number) of both the old and new images. This requires the user to make a double lookup, but this may prove to be only a minor handicap.

Roll Microfilm with Photo-optical Binary Code. This type of coding system can be used to conduct computer-like searches. Figure 14 depicts a sample of photo-optically coded roll microfilm. Such document descriptive data as titles, names, dates, numbers, and subject topics can be recorded in photo-optical binary code format on the film, thus permitting the user to automatically conduct both simple and complex or coordinate-type searches. Depending upon the features of the particular equipment, search entry is made through a keyboard, dials, or a machine record such as edge-punched cards.

The major advantage of the motorized roll microfilm system with photo-optical binary code is that it permits the user, while conducting the computer-like search, to simultaneously see the documents involved.

The major disadvantage of these systems is the cost. Except where COM equipment is employed for preparing the microfilm, the input costs are usually greater. The retrieval equipment costs more than that used in most other microfilm systems and is somewhat more difficult to operate. Unlike computer systems, the binary optical code, once recorded on the film, cannot be changed. Further, unless the file can be broken down into separate autonomous groups and the individual searches confined to a single group, the time required to conduct individual searches will increase as the file grows. This could result in a need for additional equipment and personnel, and thereby tend to offset the initial advantages of the system.

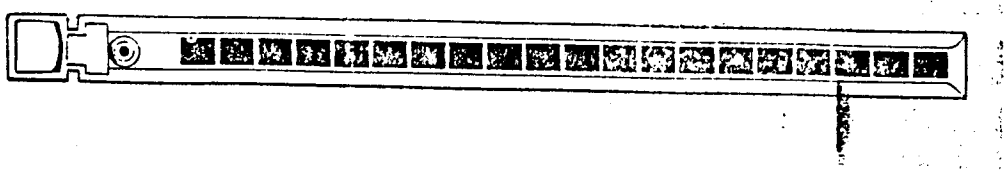
Microfilm Strip Systems. Microfilm strip systems employ roll microfilm cut into segments for storage of multipage documents. Three general manual methods used for storage and retrieval of the strips are: (1) maintenance in separate small metal or plastic containers; (2) attachment of the strip to the edge of a card or sheet containing full-size written information; and (3) attachment of the strips to plastic sticks about a foot long maintained in horizontal racks for rapid removal and refiling. The first two have received limited use for dissemination, storage, and on-demand reproduction of lengthy documents, while the third has been used primarily for storing and retrieving information and data contained in such listings as a directory or catalog. Figure 15 depicts a microfilm strip attached to a plastic stick, and the special storage rack and reader used for this type of strip system.

All three techniques provide a means for unitizing microfilm so that the individual documents or parts thereof may be independently selected, viewed or copied, and refiled. The third technique facilitates storage and retrieval of lengthy listings by making it possible to keep them in a very small space while at the same time permitting random, fast access to the information. However, an actual test is always needed to determine comparative retrieval speeds.

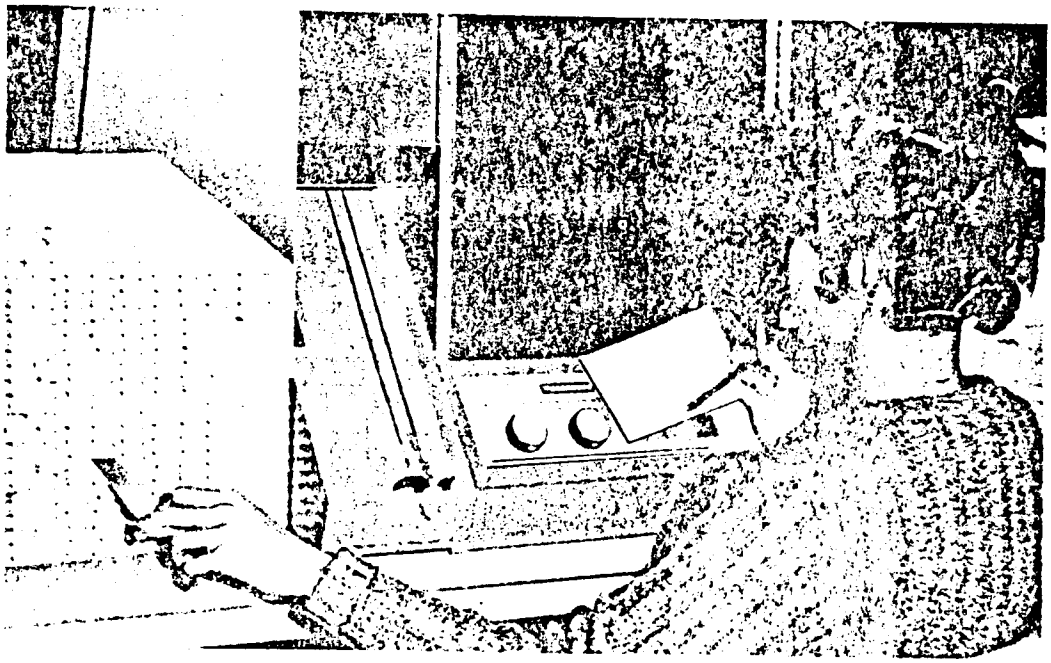
The major problem with the first type is that of physically handling the strips—opening the container, hand threading it through a reader or splicing it onto another length of film, and returning it to storage. The main problem with the second type is that it, too, is somewhat awkward to handle and can only be used in certain microfilm readers. The main limitation of the third type is the cost of preparing and mounting the film and purchasing the special reader required to view the film.

Microfilm Chip Automated Systems. These systems, as illustrated in figure 16, usually employ small pieces of cut microfilm that are often stored in cartridges or cells and manipulated by means of electronic circuitry and electromechanical devices. A keyboard or other device is used to conduct searches. These systems have been used primarily to meet the need for high-speed retrieval of short documents (one to three pages, generally) from extremely large files.

In some systems, a considerable amount of photo-optical binary coded data can be entered on the chip, while in others only a document number or address can be recorded. In one system



The Microfilm Stick



Special Storage Rack and Reader

Figure 15

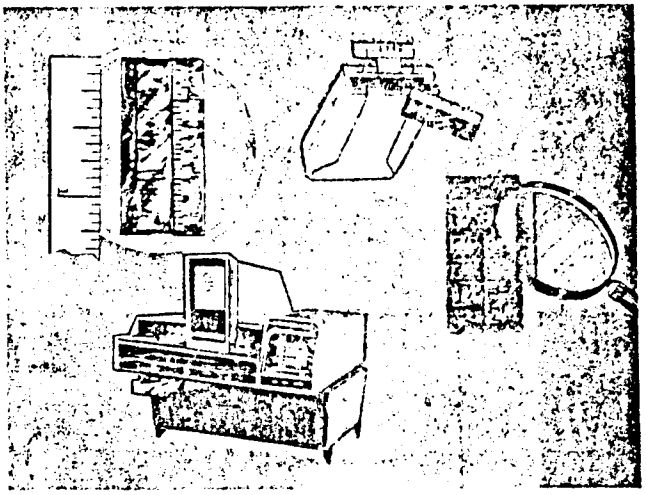
there is an iron oxide coated strip for recording data by means of a magnetic binary code, as on the magnetic tape used on computers. Microfilm chip systems are quite complex, usually involving rather high equipment costs, and thus have not been used as extensively as some of the other systems. The hardware is generally not available off the shelf but must be custom engineered.

Microfiche. Microfiche, as illustrated in figure 17, are sheets of microfilm containing multiple rows of micro-images arranged in a grid pattern. Microfiche are particularly well suited to the reproduction, dissemination, storage, and retrieval of documents or records having a total length of 20–98 pages or having chapters, sections, or parts of that length; they can also be used for longer documents, of course. Microfiche are sometimes used for storage of case-type material, such as hospital records.

The two most commonly used microfiche formats are both about 4- by 6-inches in size. The formats shown in figure 17 (60 pages per microfiche) was adopted in 1965 as the Government standard for reproduction of scientific and tech-

AUTOMATED MICROFILM CHIP SYSTEM

Figure 16



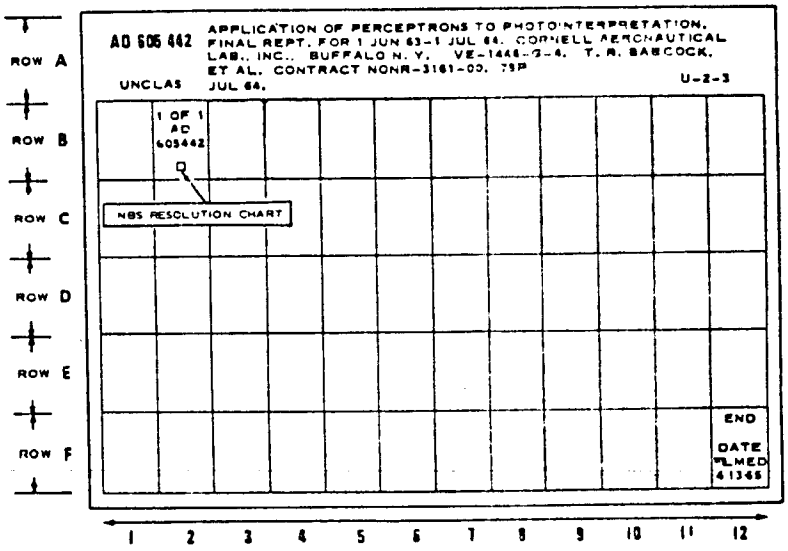
nical documents. Another format (98 pages per microfiche) has recently been increasingly used by both industry and Government. Figure 18 describes some of the wide variety of microfiche formats and reproduction ratios in use today, including high reduction (HR) ratios.

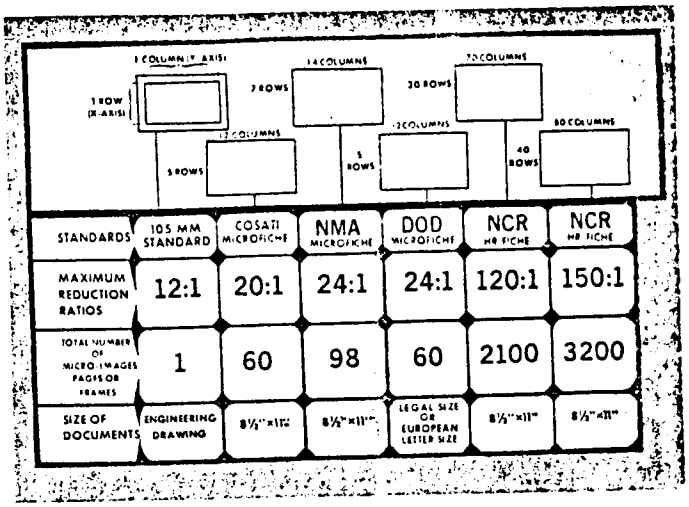
One of the major advantages of the microfiche is a possible savings of 70 percent or more to the user in acquisition costs in instances where a document is available in both microfiche and paper form. Another advantage is the elimination of document warehousing problems, since low-cost copies of microfiche can be produced at any point on demand. In many situations the most significant advantage is the savings in time and costs for packaging, shipping, storing, and retrieving documents.

Probably the major disadvantage of the microfiche is the relatively high input cost, which may make this type of microform uneconomical for internal application within a single office. However, if the documents are widely distributed, input costs can become quite insignificant. Another disadvantage is that there has been no prac-

SAMPLE OF A MICROFICHE

Figure 17





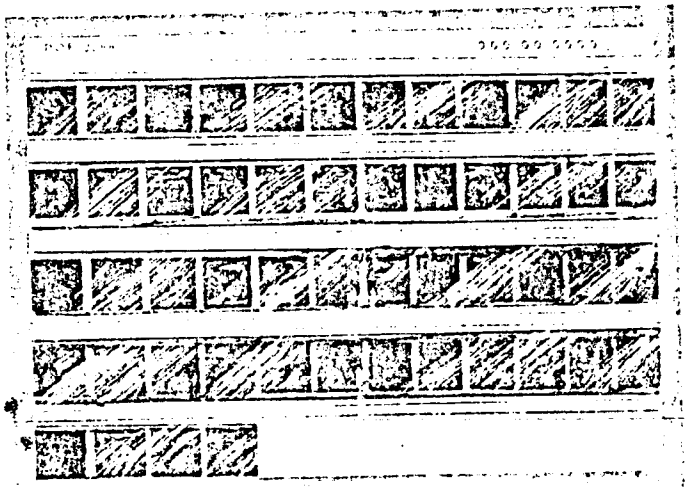
COMMON MICROFICHE FORMATS AND REDUCTION RATIOS

Figure 18

tical, inexpensive method for changing or adding to individual microfiche (up to the time this handbook was prepared). If updating is required, the alternatives are similar to those described above under the heading, "Special Note on Changing or Adding to Roll Microfilm."

Still another factor limiting the use of the microfiche is that special readers are required at every point of use; and even though inexpensive readers are available, the overall equipment investment may be substantial. However, as the use of the microfiche is extended to more and more document series, the readers may eventually become standard office equipment. Another possible disadvantage is that some users feel that further improvements are needed in the readers in order to make the viewing more convenient and comfortable.

Microfilm Jackets. Microfilm jackets are transparent carriers with one or more sleeves or pockets for holding strips of microfilm, as shown in figure 19. The entire jacket, with the microfilm inside, is placed in a reader for viewing. Film-to-film copies and paper enlargements may be made without removing the film from the jacket. To get the best results it is necessary to use one of the newer "thin film" jackets.



SAMPLE OF A MICROFILM JACKET

Figure 19

SAMPLE OF AN APERTURE CARD

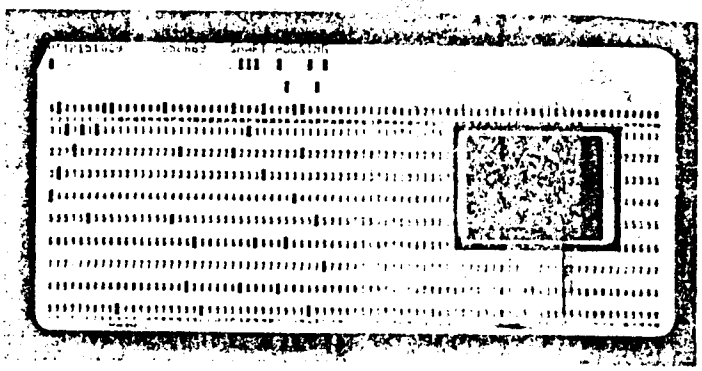


Figure 20

The major advantage of the microfilm jacket is that new images may be added, thus making it particularly suitable for active case-type records. It is compatible with the microfiche and can be used in the same types of readers and film-to-film copiers, and thus has many of the advantages noted above for the microfiche.

The major disadvantage of the film jacket is the time required for inserting individual micro-images into the sleeves of the jacket; however, special equipment has been developed for this purpose to make the task much easier.

Aperture Cards (Microfilm Electric Accounting Machine Punched Card). These cards, illustrated in figure 20, are standard punched cards (or edge-notched cards) with windows containing micro-images. The window is usually designed to accommodate one large document, such as an engineering drawing, or as many as eight or 10 letter-size pages, which in the case of the punched card, would require 22 card columns of space. This leaves over 50 columns for recording data such as the document number, description, and date in machine-coded form. There are also aperture cards containing sleeves as in microfilm jackets for inserting and adding images.

One of the major advantages of the aperture card is the convenience in filing, retrieving, and adding to the file. Another advantage of aperture card systems is the capability for using mechanical devices for sorting and selecting individual cards, while at the same time permitting manual filing and selection of cards. Still another important advantage is the savings in time and cost for duplicating, shipping, handling, storing, and retrieving documents. Further, there is available a

SUPERMINIATURE (HIGH REDUCTION) MICROFORMS

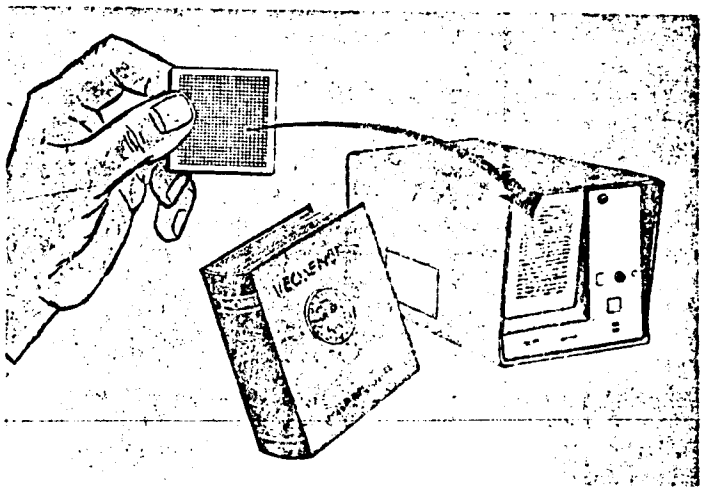


Figure 21

VIDEO RECORDING SYSTEM

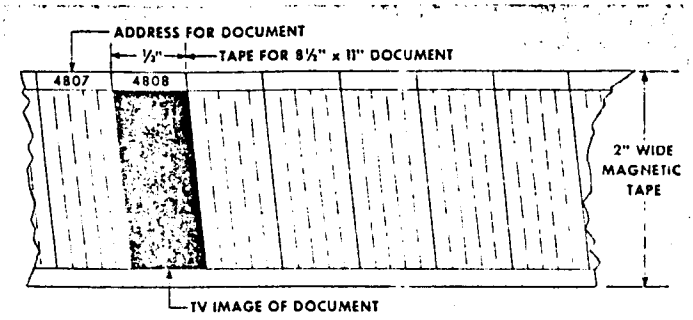
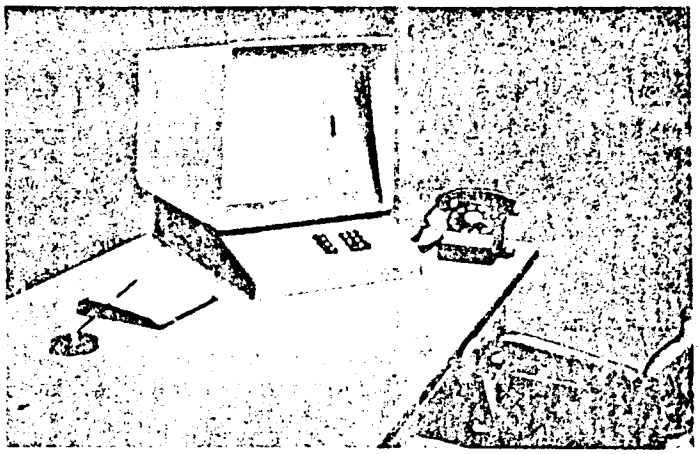


Image Storage



Desktop Viewing Equipment

Figure 22

wide variety of equipment to satisfy the needs of the smallest to the largest user.

The major disadvantage of the aperture card system is the relatively high input cost involved in the filming, keypunching (or edge-notching) of the cards, and mounting of the micro-images in the apertures. Therefore, as in the case of microfiche, the cost may make such systems uneconomical for internal application within a single office. Further, extensive machine sorting and selection of the aperture cards may not be practical if the file is a very large one. When punched card equipment is used for card sorting and selecting, it is usually modified in order to minimize damage to the micro-images; or a duplicate "slave" deck, which does not contain the micro-images, is created for use in the punched card machines.

Superminiature (high reduction) Microforms. Superminiature microforms and those referred to as ultraminiature microforms (ultrafiche) employ a reduction ratio much higher than those used for ordinary microforms. (See figure 21 for an example of a book of more than 1,000 pages reduced to one ultrafiche, and the special reader required for viewing the images.) The standard reduction ratios in use today readily permit the recording of 2,000 to 2,500 letter-size pages on a 100-foot roll of microfilm (and in some systems, up to 4,000 pages per 100-foot roll). Reduction ratios as low as 10 to 1 (10X) are used for newspapers and as high as 42 to 1 (42X) are used for COM produced listings and cancelled checks. Superminiature microfilm, on the other hand, employs reduction ratios of approximately 200 to 1 (200X) and higher.

The major advantage of superminiature microforms is the further savings in space and shipping costs resulting from the greater compactness of the micro-images. Superminiature microforms make it possible to store an extremely large collection of documents close to the users, or possibly within the viewing equipment itself.

The major disadvantage of the superminiature microforms is the initial cost of preparing the master copy. However, as in the case of the microfiche and the aperture card, this cost may not prove excessive if there are a large number of users at various locations who use the same information over and over again. Another disadvantage is the lack of compatibility between this and any other microform media. Special readers with optics compatible with the very high photographic reductions are required.

Video Recording Systems. These systems employ the basic techniques and equipment used in recording television programs, as illustrated in figure 22. Documents are placed under a camera and magnetically recorded on video tape or other media. There is a separate track for recording the document's number or other identifier. Retrieval is accomplished through a keyboard or by preparing a machine record such as a punched card that is fed into the retrieval device. Images of the retrieval documents may be viewed on remote terminal cathode ray tube (CRT) screens, or enlarged paper copies can be produced.

The major advantages of the video recording systems are the instant recording and inspection of document images; the ability to add or delete documents; the ease of use; and the relatively fast retrieval speeds. Video recording systems have not been in use long enough to fully evaluate their performance and potential. However, the major disadvantages appear to be the relatively high systems cost; the need for special skills in planning, operating, and maintaining the system; and the need for special work procedures and routines to compensate for the lack of a practical means for gaining random access to the file.

Special note on mechanized devices (miscellaneous card selectors) for storage and selection of microfiche, microfilm jackets, aperture cards, and other unit records.

There are numerous electromechanical devices that permit selection of individual unitized microforms by means of a keyboard. The smaller ones have trays holding approximately 1,000 items each, which can be interconnected and operated through a single keyboard. Typically, the individual items are notched along the bottom edge, and the selected item pops up when its identifying number or location address is entered on the keyboard.

There are also very large units, some of which can be accessed through remote terminals equipped with keyboards and CRT displays. Some also have the ability to perform limited coordinate-type searches.

The major advantages of these devices are that they reduce physical strain, eliminate the need for interfiling as microforms are returned to the file, and make possible an increase in retrieval speed.

The major disadvantage is cost. To justify the purchase of such equipment the file must be very active, but not more so than one person per keyboard could handle. Thus, the limited access to the file could pose a serious problem in times of peakloads, expanded reference activity, or machine breakdown.

Microform—Computer Combinations

The motorized roll microfilm systems with photo-optical binary code and the microfilm chip systems combine in a single medium both machine-readable data and document images for simultaneous searching and viewing of the micro-images. Further, it is possible to use any of the various types of microform methods and equipment described earlier in combination with a computer. There are, however, an increasing number of microform devices specially designed for direct use with the computer.

Computers, as explained in chapter V, can perform complex coordinate and other types of logical searches, as well as other forms of data manipulation, at fantastically high speeds. However, storage of very large volumes of data on-line can be extremely expensive; and since computers can only work with information that has been converted to a machine language code format,

their capability for storage and presentation of graphics and large masses of data is rather limited. The situation is much the reverse for microforms, of course. Consequently, the computer and the microform can often be used to complement each other very effectively by maintaining the low-volume index data (or dynamic data) on-line with the computer and the large volume of information and graphics (or static data) in microform. Finally, a communications link—either human or part human and part machine—or all machine—is provided to permit the two to work as a team.

Roll microfilm and various forms of unitized microfilm such as microfiche, microfilm jackets, aperture cards, and chips are often employed. In any event, the microform portion of the system's work station includes a microfilm reader or copier that is mechanized to some degree. Communication with the computer portion of the system may be accomplished by either of two methods: One uses a remote terminal with a keyboard and possibly a CRT display: the other uses a punched card or punched paper tape equipment for sending messages to and from the computer. There is also equipment available that permits use of a single keyboard to communicate with both the microfilm and computer portions of the system. It employs a split viewing screen for simultaneously displaying information produced by both parts of the system.

If a person serves as the communication link between the computer and the microform storage unit, he is responsible for retrieving the appropriate micro-images upon receipt of the message from the computer. In other systems the computer message is used to automatically activate a microform reader that finds and displays the related micro-images for the user. In still another system, the computer message is used to control a mechanism that locates the proper microfilm image and makes a film-to-film copy of it.

The advantages of combined microform computer systems include an increase in the usefulness of the computer, reduction of computer storage costs, faster retrieval of information, and improved access to information. By using microforms to store close at hand large masses of previously acquired information along with current static or semistatic data and then using the computer to quickly identify the location of n information and perform related ADP opera new solutions are provided for both today' tomorrow's problems.

The disadvantages are mainly that suc tems usually require highly skilled designe a rather substantial initial investment.

Special Considerations

It should be quite clear by now that micr systems do not offer a panacea for all agency's document dissemination problen cost-benefit study should always be mad pilot tests conducted before deciding to go with a system. A major obstacle in any micr system is gaining user acceptance, and nc should be left to chance. Appendix "E," partment of the Air Force Regulation 1 March 5, 1971) provides a good example types of management controls required to i the successful application of document mini zation techniques.

When designing a microform system, s consideration should also be given to capt and maintaining key identifying data in ma language. Using source data automation niques, this can be done for a small addi cost at the same time the labels are typed machine-language record should prove h useful as a means for automatic preparati finding aids, inventory lists, and new labels purging of the file.

Attention should also be given to subpar 11.5 of the Federal Property Management lations (41 CFR 101–11.5). While this regu primarily applies to situations involving filming of permanent records in order that can be destroyed, many of the safeguards vided therein should be observed in all micr systems.

The National Archives and Records S General Services Administration, operates filming service centers throughout the co Government officials interested in these se or desiring assistance in microfilming and paperwork management matters should c the manager of the nearest GSA Regional or Federal Records Center.

IV. MANUAL NONCONVENTIONAL INDEXING SYSTEMS

The methods and equipment described in chapter III, "Microform Systems," were developed primarily to solve the problems associated with the physical handling and storage of documents. It was also explained in that chapter how microform systems can sometimes prove helpful in solving problems involved in looking up data in such voluminous listings as payrolls, directories, schedules, and price lists. If, for example, the user's problem is simply to look up the social security number, address, or telephone number of individuals with whom he deals, a microform system, or perhaps a conventional tool such as a printed directory or card file, is usually all that is needed.

If, on the other hand, retrieval involves searching for documents or information on the basis of subject topics or a variety of characteristics, attributes, or other features, the problem is quite a different one. The problems and limitations in using conventional methods and equipment in situations of this type are described in chapter I, "Why Information Retrieval Systems Are Needed." Chapter I, as well as chapter II, "How Coordinate Indexing Systems Work," explains how the nonconventional information retrieval systems may be employed to solve these problems. This chapter (IV) and the next one (V) describe the specific methods and equipment used in these nonconventional systems. This chapter covers manual methods and equipment, while the one that follows describes those employing mechanized equipment.

Manual nonconventional indexing systems, for the purpose of this handbook, include those where the search is conducted by manual methods. The tool or device may have been prepared manually, but some are, and most could be produced and updated by computers and other machines. Further, some of the tools could be converted to a microform format for ease in duplication and dissemination.

Types of Situations Where Nonconventional Indexing Systems Are Used

There are two basic types of situations where the methods and equipment in this and the next chapter are applicable. The first type involves organization of information mainly on the basis of *subject topics* for retrieval of textual documents or information. The second type is concerned with organizing information (data) on the basis of *characteristics or attributes* (also referred to as indexing terms in this handbook) for use in identifying and retrieving information or documents relating to individual people, places, or things. An example of this second type is a personnel skills inventory describing employees in terms of their education, experience, languages spoken, etc., for use in selecting people for promotion, reassignment, special projects, or other purposes. This second type of system is far less complex to design and operate than the first, mainly because it is relatively simple to develop and define the characteristics, attributes, or features to be used as indexing terms, while the task of selecting and defining subject topics is difficult and imperfect due to the ambiguity of the human language.

Prerequisites for a Successful Manual Nonconventional Indexing System

The most important prerequisite for a successful indexing system is *to obtain the right people for the job.* In all but the smallest and simplest of systems, special talents of two types are required. The first requirement is for the services of a skilled person to design the system and then return periodically to revise it, since there is no such thing as a finished design for an indexing system. If the system involves indexing documents by subject, the individual should have a thorough knowledge of both the subject matter field and indexing. If no such person is available, it may be necessary to use the team approach: that is, to bring together an individual who has a thorough

knowledge of indexing but only a limited knowledge of the subject matter with a person of the opposite qualifications.

The second type of talent needed is *qualified personnel to operate the system.* Again, if the system is used for indexing documents by subject, the indexers and searchers or indexer-searchers (and abstracters, if any) should have a thorough knowledge of the subject matter field and be properly trained in performing their duties.

Of next or perhaps equal importance is the need for *an operating manual or rule book.* The operating manual should include a vocabulary of indexing terms or a thesaurus, as it is commonly called, listing all indexing terms and defining how they are used in the system, supplemented by cross-references for synonyms and incorporating one or more devices for showing relationships among indexing terms. The operating manual should also include any other rules, guidelines, and reference aids needed for indexers, searchers, and users.

Another prerequisite for a successful system is *close coordination between the operators and users of the system* in all matters, including selection of documents or data entered into the system and continuous feedback on the effectiveness and value of the system. All users need to be kept informed about the new accessions, and new users should be oriented in regard to the contents and use of the system.

Another possible prerequisite, or at least desired feature of the system, is *compatibility with other systems* with which it may be interfaced now or in the future. This compatibility is of two kinds—system vocabulary and physical aspects. Today, it is seldom that any given collection of documents or data is of interest or value to a single organizational element. Somewhere within an agency, another agency, or the private sector, there is likely to be one or more groups of people collecting, storing, and retrieving similar if not identical information. System compatibility can therefore be of mutual benefit, possibly contributing through sharing arrangements to lower costs for all systems involved, while increasing the level of service to users.

Another important prerequisite is that there should be *a minimum of delay in entering the i items into the system* and making them availa to the users. Not only should a search of the in reveal the presence of the item, but it should a be possible for the user to quickly obtain a co of it.

Other prerequisites for providing good serv to the user include ready access to the system a satisfactory performance of the system. A hig desirable but not necessarily essential feat would be that the system be readily converti to an automated system.

Factors Affecting the Choice of the Ty of Manual Nonconventional Indexing System

The major factors to be considered in choos the most suitable type of manual nonconv tional indexing system are as follows: (1) present file size, growth rate, and estimated ture size of the collection; (2) if the system is be used for retrieving information by subject, average number of indexing terms that will assigned each document and the total numbe indexing terms for the system; (3) if the syst is to be used for retrieving information or id tifying people, places, or things on the basis characteristics, attributes, or features, the num that will be used to describe each item ente into the system: (4) physical form, format, c and source of the input; and (5) the extent which the documents or data will have to changed, updated, or deleted.

Other important factors to be considered selecting the type of system include: (1) the av age number of indexing terms to be used search, the average number of searches per d and the extent of workload fluctuations and pe loads; (2) the number and types of users a their physical location; (3) the physical fo format, and nature of the output required by users; (4) service speed requirements; (5) spe features required, if any, such as abstracting evaluating documents and selective dissemi tion of information (SDI); (6) accuracy and liability requirements; and (7) agency resou including availability of funds, personnel, e equipment for operation of the system.

Further information regarding the significa

CLUE-WORD EXTRACT CARD SYSTEM

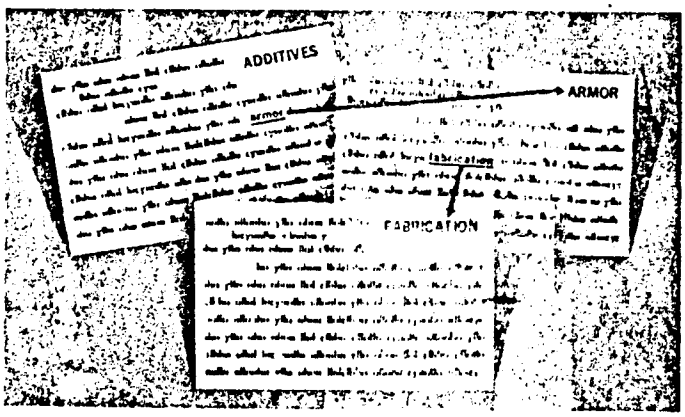


Figure 23

of these factors and guidelines on gathering the data, analyzing user needs, and selecting the right method and equipment are included in chapters VI and VII. A description of a number of systems employing manual nonconventional indexing methods and equipment is included in the records management handbook, *Information Retrieval Systems.*

Types of Manual Nonconventional Indexing Systems

The following are descriptions of the various types of manual nonconventional indexing systems, with a brief summary of the main advantages and disadvantages or limitations of each.

Clue-word Extract Card Systems. These systems are subject indexes consisting of 5- by 8-inch cards arranged alphabetically by "clue-words" (keywords) taken from the titles and text of the documents. Each card contains an extract of the document in which the keyword appeared. The extract is marked to indicate other keywords contained in the document, thus providing built-in "clues" as to other places to look in the file when conducting a search. Figure 23 illustrates how the "clue-word" principle operates. Information specialists, or preferably users of the system, evaluate incoming documents for relevancy. They underline the keywords in each selected document and place brackets around the portions to be extracted. They also assign additional indexing terms, if needed. Typically, tables of contents, author-prepared abstracts, and key illustrations are included in the extract.

Typists then prepare a 5- by 8-inch duplicating master containing the document number, title, author, other standard descriptive headings, and the extract with all keywords underscored. A sufficient number of cards are made of each document to permit the filing of one card under each of its keywords and the standard headings. Various colored cards, colored stripes, and corner cuts are employed to code the cards as to date, source, type of document, etc. The incoming material is maintained in a separate file.

The user begins his search by choosing a keyword he thinks should be helpful in identifying documents that may contain the information he is seeking. If, after scanning the cards filed under that particular term, he is still unable to find what he wants or needs further information, he takes note of other underlined keywords appearing in the body of the cards for "clues" as to where else to search for the needed information. He then refers to the other cards and thus proceeds with the search until he finds the desired information or until he has satisfied himself that the document collection contains nothing significant on the subject.

The major advantages of the clue-word extract card system are that no complicated input and output equipment is required; no preconstructed index vocabulary is needed (system is self-organizing); no special training is needed for conducting searches; it is highly browsable; and the extract cards are self-sufficient (it is usually not necessary to refer to original document). Further, this technique offers a simple, effective means for compacting text. The system concept is

SAMPLE PAGE FROM A PERMUTED (KWIC) INDEX

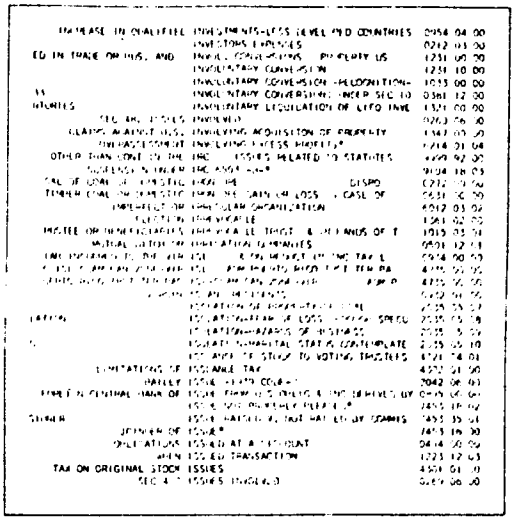


Figure 24

susceptible to application of computer techniques for information dissemination, automatic searching, and preparation of special finding aids. Therefore, anyone establishing a manual clue-word extract card system today should capture and retain the input data in machine language format for possible conversion to a computerized system at a later date.

The major disadvantages of the original clue-word extract card systems are the bulkiness of the files and the lack of a practical means for converting the file to an automated system.

Permuted Indexes. Permuted indexes are specially printed and organized printed manual indexes usually prepared by a computer from document titles, full text, a catalog, or index entries, as illustrated by the format of the KWIC (keyword-in-context) index shown in figure 24. There are various other formats, many of which are an improvement over this one. Some of the better known other permuted indexes are KWOC (Keyword Out of Context), WADEX (Word and Author Index), and SPINDEX (Special Permuted Index). To obtain these indexes a computer is programmed to alphabetically arrange the entries so that each document or other thing being described in the index is listed under each of its keywords.

KWIC indexes have been successfully ap in indexing operating procedures and direct forms catalogs, the Controller General's sions, and in numerous other situations. In where permuted indexes are used for ind procedures and directives, a special dividend be expected—the index will highlight inconsi cies, duplications, and omissions. With th creased usage of permuted title indexing, au are giving more attention to selecting meanin useful titles; and this, together with the impr formats and low costs, is enhancing the u permuted title indexing. The retrieval capab of permuted indexes can be increased by i sion of additional indexing terms selected an index vocabulary such as the *Thesaur Engineering and Scientific Terms* used by Department of Defense and other Governm agencies.

The major advantages of permuted ind are the following: (1) the relatively low ov cost (in some situations the index can serve low cost substitute for manually prepared ind or can make it practical to provide an index w none existed before); (2) speed and ease of p aration (computer printouts that serve as copy for offset printing of the index can be tained in a matter of hours); (3) ease of revi (the speed of a computer makes it possibl print out an entire new index including any r sions, rather than trying to manually patch printed copy as revisions are made); (4) r meaningful and browsable than convention printed indexes (the one or two word entrie not normally provide an entire concept); and reduction in the time required to announce documents and enter them into the system.

Permuted indexes can also serve as a m for developing in-house capability in the us computers for information processing, and i least some instances will result in the estab ment of a computer data base that may s even more important purposes in the future.

The major disadvantage of the permute dex is that it does not provide cross-reference synonyms; therefore, it is subject to searc problems created by the author's inconsiste in word usage and the normal ambiguity o man language. Further, if limited to docu titles only it becomes a shallow index; if ap

the entire text, it may become too cumbersome to be practical.

An ideal permuted index for procedural manuals and similar publications would include a permuted listing of titles for the parts, chapters, sections, the paragraph or other headings, and any abstracts or other summaries of the contents of the documents.

Columnar Card Systems. These systems, as shown in figure 25, are coordinate indexes in which one card is prepared for each indexing term used in the system. The numbers of all documents indexed under each term are entered on its term card. Each term card is divided into ten columns, 0 through 9, and the document number is posted in the column corresponding to its terminal digit. Searches are conducted by selecting those term cards that seem pertinent, and then matching document numbers column by column to locate any numbers common to all the cards. The cards are usually prepared and maintained manually, either by hand or typewriter; the basic data, however, could be maintained in machine language form and the cards produced by a computer.

The major advantages of the columnar card systems are that the costs for supplies and equipment are extremely low; they permit parallel searching of the index file (rather than requiring a card-by-card serial search); and they are simple and easy to maintain and use, being highly manipulative and browsable.

The major disadvantages or limitations of the columnar card systems are that it is usually necessary to refer to a second document, such as an abstract or even the document itself, to obtain a document description or to determine a document's relevancy; and if the system is used extensively, searching can become slow and tedious should the columns of numbers become long and individual searches involve several indexing terms.

Dual Dictionary Systems. These systems, illustrated in figure 26, are similar in design and use to columnar card systems, except that all the indexing terms and document numbers are printed on two identical lists mounted side by side in a binder. Instead of matching cards during the search process, the user looks up the first term

SEARCHING WITH COLUMNAR CARDS

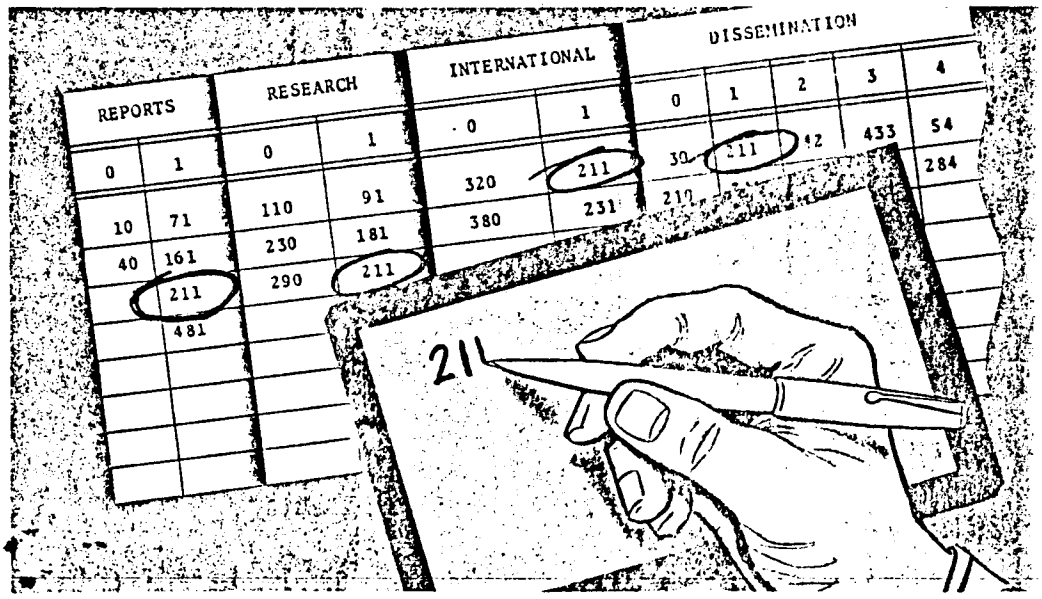


Figure 25

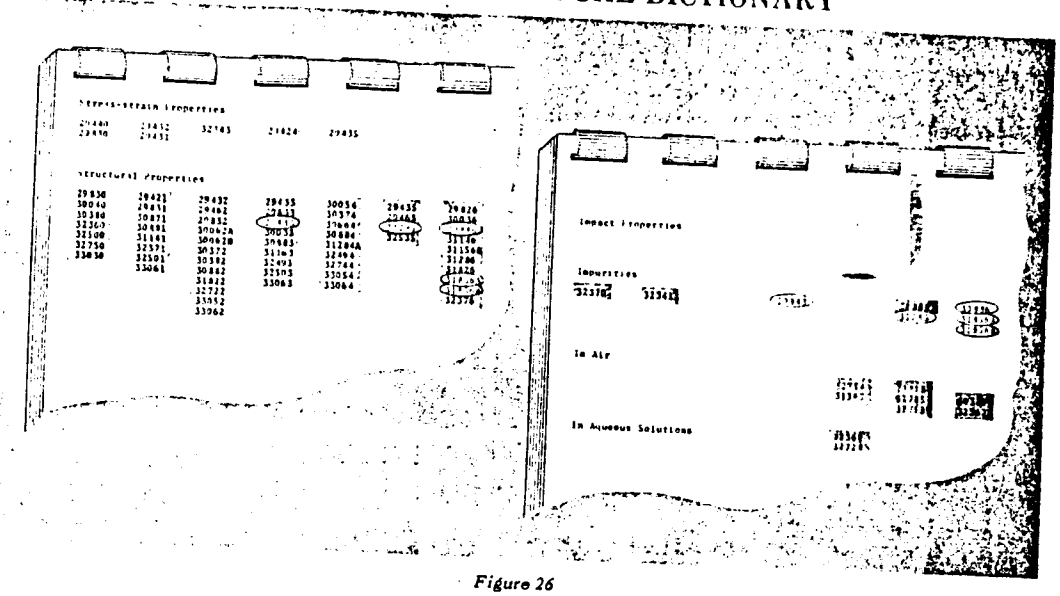


Figure 26

in its alphabetical location on the left side of the dual dictionary and then locates the second and other terms on the right side (or vice versa), checking for coinciding numbers at each step until the search is completed. Usually many copies of the dual dictionary are made and distributed to individual users.

The dual dictionary is best suited to those situations where there are many users in different locations. The dictionary's usefulness can be increased by furnishing with it abstracts of the documents and a copy of the thesaurus or other vocabulary of indexing terms.

The data for the dual dictionary may be manually maintained; however, more often it is maintained and updated by computer and then periodically printed out, duplicated, and distributed to the users.

The major advantages of dual dictionary systems are the same as those for the columnar card system, plus an important, additional one—these systems permit numerous individual users or groups of users to do their own searching, thus reducing the workload at the main information center and giving the user direct access to the system.

The disadvantages of the dual dictionary are also essentially the same as those for the columnar cards, with one exception—they are far more costly to maintain; however, if the number of users is sufficient the overall systems costs could, by comparison, be relatively low.

Edge-notched Card Systems. These systems, as illustrated in figure 27, are cards containing punching positions, represented by pilot holes along one or more of their edges, used in recording in coded form such data as indexing terms, dates, and numbers. The data is recorded by punching out the area in front of the pilot hole. The edge notching may be done manually by a hand punch or semiautomatically by special equipment. The interior of the cards, which are printed in various sizes and formats, may be used for written information or graphics. Typically, one card is prepared for every document or item being indexed.

To search the file, needles are passed through the appropriate pilot holes in the deck of edge-notched cards. The selected cards (those that are notched) fall out, while the others remain on the needle. Searching usually involves numerous needle passes. Other devices and equipment, in addition to the standard needles, are available for assisting in the search process.

EDGE-NOTCHED CARDS

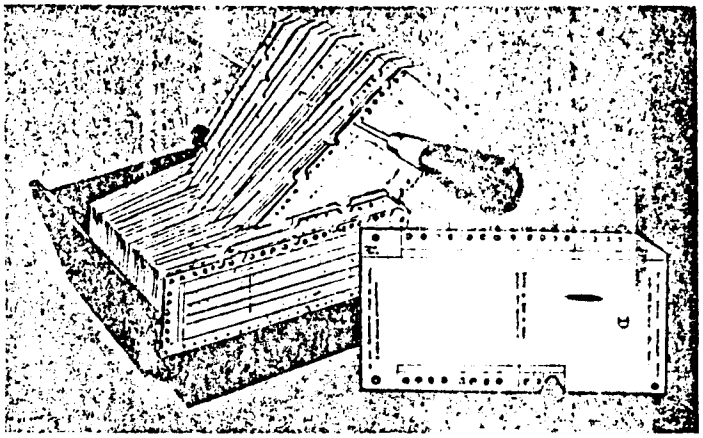


Figure 27

The major advantages of the edge-notched card systems include low cost, simplicity, the ease with which users may browse, immediate access to the description of the documents or things involved in the search process, and in many situations, elimination of the need to maintain the cards in a precise sequence.

The major disadvantages of the edge-notched card systems include limitations on the amount of coded data that may be recorded on the card; slowness and awkwardness in the search procedure if the cards are used extensively for complex searches (due to the system requirement of serial searching); limitations on the size of file (many information specialists consider 5,000 cards to be the upper practical limit); the somewhat complicated code patterns; and the possible difficulty in detecting coding (edge-notching) errors.

OPTICAL COINCIDENCE CARDS AND VIEWER

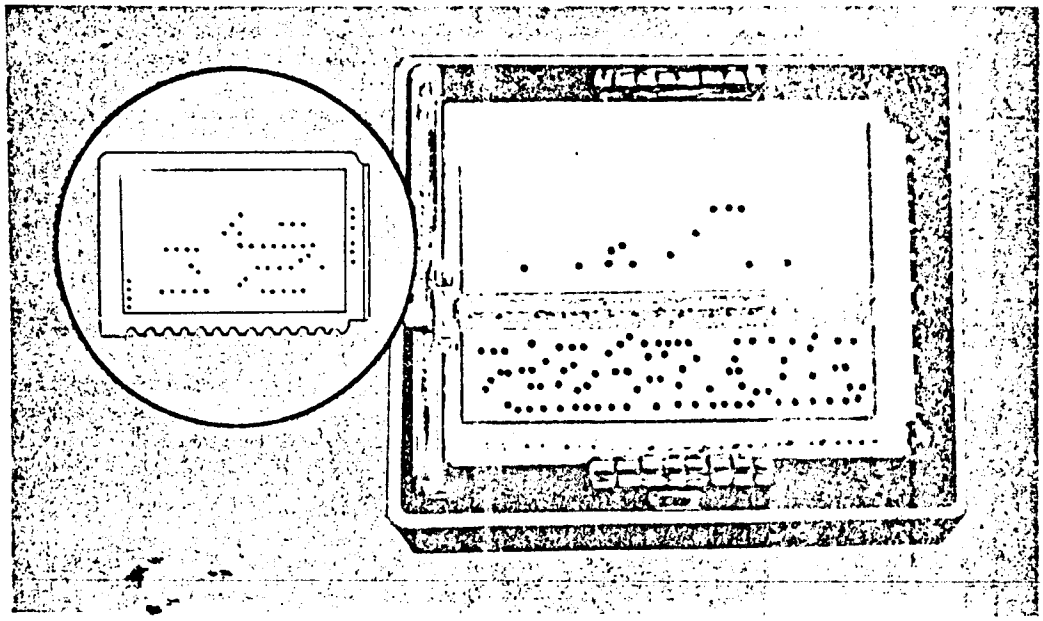


Figure 28

Optical Coincidence Systems. These systems, as illustrated in figure 28, employ cards (or sheets) with a fixed number of dedicated positions or address locations for drilling (or punching) holes representing the individual documents or items being indexed. A separate optical coincidence term card is maintained for each indexing term. After each incoming item has been indexed and assigned a serial number or optical coincidence card address location, all related term cards are removed from the file and machine drilled or punched in the appropriate position.

Searching is accomplished by first selecting the optical coincidence term cards pertinent to the query. The selected cards are then stacked and are placed in front of a light source to visualize the existence of coinciding holes. The position of the matching holes on the cards indicates the number or address locations of any documents or items that fully satisfy the search question.

In addition to identifying documents or other items pertinent to a query, the cards may also be used as a data manipulation and tallying device for compiling statistics; or, through the use of transparent overlays, as a means for presenting statistical data in a visual manner. Although in most optical coincidence systems the cards are drilled, manipulated, and interrogated manually, there is equipment available for machine controlled drilling of the cards, machine counting of holes, and automatic printout of numbers. In the system developed by the National Bureau of Standards, the user can see an enlarged microfilm image of the related document abstract during the interrogation process.

The optical coincidence cards most commonly used are about 9 inches in size and can accommodate up to 10,000 documents or items and 1,000 indexing terms. Prescored punched cards that can accommodate 480 items are also sometimes used.

The major advantages of optical coincidence systems are manipulatory ability; encouragement of browsing by the user; rapid searching speeds (partly because these systems permit parallel searching of the index file rather than requiring a serial card-by-card search); low cost for supplies and equipment; simplicity; and fast, easy read-out of the search results.

The major disadvantage of optical coincidence cards is that it is usually necessary to refer to a second information source to obtain a description of the document or item, or to determine its relevancy. Another possible problem is in error correction; however, some types of input equipment help keep errors to a minimum by preventing redrilling in the same hole.

Special Considerations

This chapter reveals that there are many simple, rather inexpensive nonconventional indexing systems which, although manually operated, offer significant advantages over conventional systems for organizing and retrieving information. In many situations today, one of these manual systems may be all that is needed to solve the information retrieval problem. However, in most situations it will some day become desirable to convert the system to one that takes advantage of computer capabilities for maintaining, reorganizing, reformatting, merging, updating, and purging of information in the file, and manipulating, selecting, and presenting the information.

In order to do these things the data contained in the index file must be in machine language. Consequently, when developing and installing any manual nonconventional indexing system, serious consideration should be given to recording the index data in machine language as a by-product of the input operations. Such devices as paper tape and magnetic tape or card typewriters are ideally suited to this purpose. Further, as mentioned earlier in this chapter, the machine language data base, with the aid of a computer, can be used to produce many of the nonconventional manual indexing tools.

V. NONCONVENTIONAL MACHINE INDEXING AND RETRIEVAL SYSTEMS

The significance of nonconventional machine indexing and retrieval systems rests not in the number of basic types of equipment that are available, but in the wide variety of tasks these systems can perform, their flexibility, and their future potential. In numerous instances the indexing, storage, and retrieval operations are, or could be, a satellite of a larger integrated automatic data processing (ADP) system. Today, there are many instances where the data base maintained for an ADP system could, with slight modification and expansion, serve as the nucleus for a highly useful information retrieval system. On the other hand, there are situations where machine nonconventional indexing and retrieval systems could largely pay for themselves by solving logistical and other problems involved in the preparation, stocking, distribution, replenishing, and control of documents.

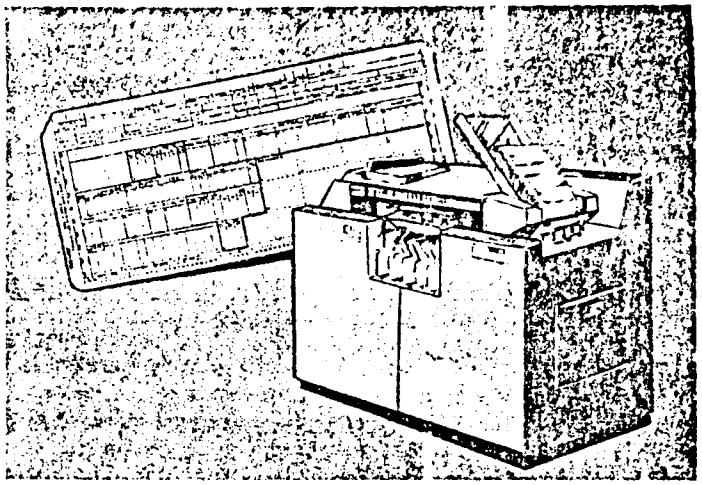
Obviously, the equipment used in machine nonconventional indexing and retrieval systems is usually more expensive than that used in the manual systems. Further, the machine systems are generally more difficult to design and operate. However, these conclusions can be misleading, and in practice they prove to be but a slight barrier in installing a machine system. The first reason for this is that instead of acquiring your own equipment, you could more than likely obtain machine time on equipment already installed in the agency or available through a service bureau. The second reason is that there are available many standard and special machine programs (machine instructions and procedures) that, with slight modifications, can be adapted to the job at hand. When one considers these possibilities, and the indisputable move toward automation in all areas, it becomes increasingly clear that any information retrieval system study should include a thorough investigation of machine methods for doing all or part of the job either now or in the future.

Types of Situations Where Machine Indexing and Retrieval Systems Apply

There are two basic situations where the methods and equipment described in chapter IV and in this chapter may apply. In the first situation, i.e., *retrieval of textual documents or information on the basis of subject topics*, machine systems are proving highly satisfactory: and in addition, many of the systems can automatically furnish the user with a complete description of the document or permit him to view the document or—perhaps immediately—to obtain a copy of it. In the second type of situation, i.e., *retrieval of information or documents on the basis of characteristics or attributes*, machine systems have the additional capability of being able to automatically retrieve selected data about a person, place, or thing, or a complete description or image of it. There is also an additional type of situation where only nonconventional machine information retrieval systems apply—*the storage and retrieval of large masses of data in what are commonly called data banks*. Machine methods and equipment can be used to update these files, to automatically and selectively transfer data from one file to another, and, on demand, to selectively retrieve data and perform data manipulations.

Prerequisites for a Successful Machine Indexing or Retrieval System

All the prerequisites cited in chapter IV for a successful manual nonconventional indexing system are also important to the success of machine systems, and therefore should be carefully noted. An additional prerequisite for machine systems is the ready availability of personnel, either on a full or part-time basis, who are trained and experienced in the operation of the equipment. Another important prerequisite is the accessibility of equipment being able to have access to it at the right time and frequency required by the users.



EAM PUNCHED CARDS AND COLLATOR

Figure 29

Another important but not necessarily essential feature is that the data elements and codes be compatible with other computer data banks in the same field of interest so if it should later become necessary or desirable the data can be readily exchanged, compared, or combined on a machine-to-machine basis.

Factors Affecting the Choice of the Type of Machine Indexing or Retrieval System

In addition to the factors cited for manual systems in chapter IV, which also apply here, machine systems are concerned with machine record lengths. Machine record lengths involve: the number of data elements (for example, date of birth) per record; the number of data items (for example, year of birth) within the data element; and the total number of characters (alphabetical, numerical, and special) per record.

Types of Machine Indexing and Retrieval Systems

The following are descriptions of the various types of machine nonconventional indexing and retrieval systems, together with a brief summary of the main advantages and disadvantages or limitations of each.

EAM (electrical accounting machine) Punched Card Systems. These systems employ cards divided into vertical columns, with each column then divided into 12 punching positions. Each column can be used to record, by means of one or more punched holes, a single alphabetical, numerical, or special character. The cards are divided into segments (fields) of various lengths for recording such individual data elements as the following: titles, segments of text, names, dates, addresses; and code numbers representing names of organizations, forms, products, or indexing terms. A wide variety of equipment is available for punching, sorting (including electronic high-speed sorters), collating, interpreting (card printing), selecting, and analyzing the punched cards, in addition to equipment for performing arithmetic operations and preparing printed listings. Figure 29 illustrates a punched card and a special collating machine.

Punched card systems were originally intended for use in performing statistical and accounting operations. In using punched cards as a medium for recording and retrieving data for information retrieval, the system designer has to adjust his methods to the capabilities and characteristics inherent in punched card equipment.

In organizing a punched card file for a coordinate index, there are two general ways for recording the index data and arranging the punched card file. One way is to prepare one or more punched cards, as needed, for each document or other thing being indexed and record thereon a

limited amount of data identifying the document, plus all the assigned indexing terms (subjects, characteristics or attributes). The file is arranged in document number sequence. The second way to organize the index file is to prepare a separate punched card for each indexing term assigned each document. Each card usually contains only the document number and the assigned index term; the cards are arranged in groups according to the indexing terms. This is commonly referred to as an inverted file.

The first way of organizing the file has the disadvantage of making it necessary to pass the entire punched card file through the equipment each time a search is conducted; however, it has the advantage of furnishing the user at least a brief description of the document. The second approach has the advantage of making it necessary to process only those punched cards representing the indexing terms involved in the search, which is conducted by comparing the punched cards representing any two of the indexing terms to determine coinciding document numbers, and repeating the matching process for the remainder of the term cards involved. This second method has the disadvantage of providing the user with the document numbers only, thus making it necessary for him to refer to a second source or to the document itself to obtain a description of the document and determine its relevance to the search question.

Another method of recording the indexing terms on the punched cards is to use superimposed coding, which offers greater data compaction but requires considerably more skill on the part of the system designers and operators.

The major advantages of punched card systems, when used for information retrieval, are their ease of manipulation; their relative simplicity (when compared with computers); their ease in reformatting, transferring, extracting, updating, and duplicating data; their capability for producing low-cost duplicate sets and printed listings; the ability of the cards to also be manually selected, read, and refiled; and their ready convertibility to computer systems.

The two major disadvantages of punched cards used as information retrieval systems are (1) the relatively slow searching speeds and the resulting slowness of the entire process of conducting coordinate-type searches; and (2) the limited accessibility of the punched card system, including a restriction upon the freedom of the user to browse, due to the fact that card files and equipment are usually maintained in a machine room and their use requires trained machine operators.

Most systems employing punched cards for coordinate indexing consist of less than 20,000 cards; however, if used primarily for simple data lookups and only occasionally for coordinate-type searches, a file of 50,000 or more may be feasible. Therefore, punched cards, due to this reason and the advantages described above, are particularly well suited to personnel skills inventory and other systems that usually entail a large volume of manual data lookups and recurring or special printed listings of various types and formats, but only a limited number of coordinate-type searches. Punched cards may also be used for selective dissemination of information (SDI) systems, but since today computers are more often used for this purpose, selective dissemination of information systems are included in the latter category.

Computers. Computer equipment is of two basic types: analog and digital. Analog computers may be likened to a slide rule or an automobile odometer, since they work with physical quantities and compute by measuring. Digital computers, on the other hand, work with numbers or digits and compute by counting. Digital computers are divided into two classes, special and general; computers in the general class are normally used for automatic data processing (ADP) and information retrieval. A typical equipment configuration is shown in figure 30.

Computers are the most versatile and powerful of all the devices used for information retrieval, due to their high processing speeds, accuracy, ease of updating, ability to perform complex transactions automatically and to communicate with each other, and their ability to provide the user with a wide range of on-line search capability and off-line services and tools, including permuted indexes such as the KWIC index described in chapter III. Another advantage offered by the computer used for information retrieval purposes is its usefulness for administrative and

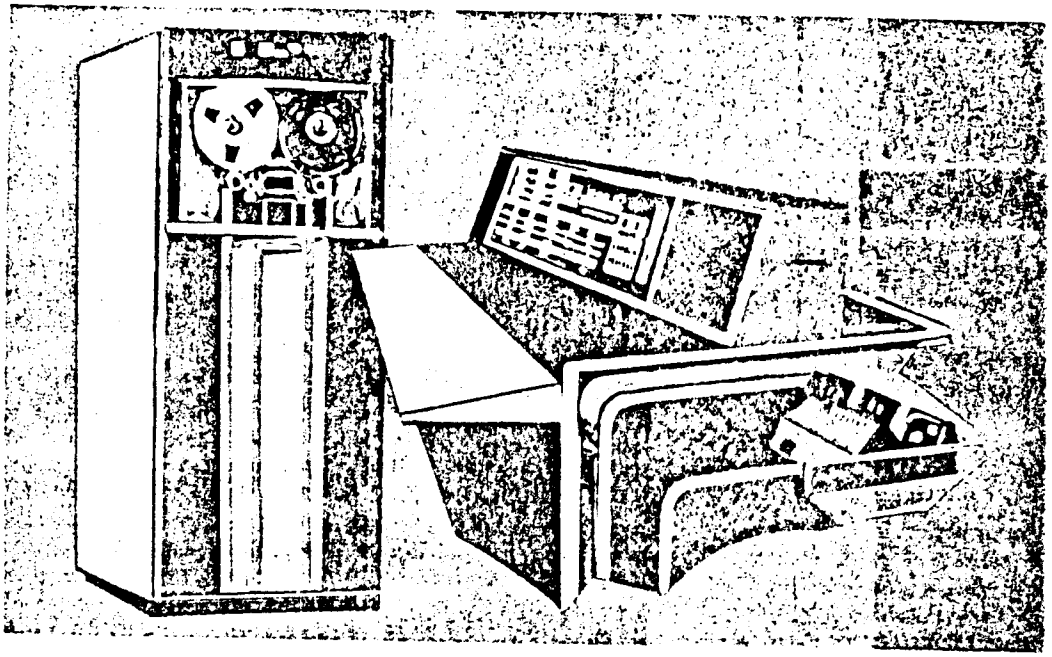


Figure 30

logistical tasks. For example, it can be used to prepare requisitions and announcements of new accessions, to operate a selective dissemination of information system (SDI), to bill for user charges, and to maintain an inventory. These special tasks are all accomplished as a by-product of normal input and output operations. The computer can also be useful in controlling access to restricted or classified information.

Three of the major limitations in using the computer for information retrieval are (1) high input costs; (2) shortage of systems analysts and programmers having experience in information retrieval systems; and (3) lack of low cost, on-line computer mass memories. Solution to the input problem depends on applying source data automation (SDA) techniques, including capturing data in machine language as a by-product of other processing operations and using optical character recognition (OCR) equipment for automatic document reading and conversion to machine language.

The problem of the scarcity of experienced systems analysts and programmers in the information retrieval area is still acute, and the only significant relief available at present is to utilize existing computer programs and operating systems developed and designed by others. The problem of developing low cost, on-line mass memories is the object of intensive research by many computer manufacturers and others, and while the results look promising, none are yet commonly available.

Notable progress has been made in computer-user communications. While most systems still require the preparation of a punched card to gain access to the computer and most of the output is still in the form of printed forms and listings, punched cards, or microfilm produced by COM equipment, there are more and more systems that permit direct communication between the user and the computer.

These two-way communications are accomplished by means of remote terminals employing teletypewriters, other types of typewriters, and cathode ray tube (CRT) devices with keyboards and light pens, as illustrated in figure 31. By keying in the proper user identification code and

following a prescribed routine for interrogating the computer, the user is able to obtain answers to his questions or possibly update, edit, or delete data in the computer store. With the addition of the light pen, he is able to pinpoint numbers, words, or phrases appearing on a CRT to make searching easier and faster or to quickly instruct the computer to delete, change, edit, or transfer stored data. Data in the computer store can also be used to produce charts and other graphics.

Significant refinements in computer programs, which make communication with the computer more like conversation, plus improvements in the hardware and reduction in equipment costs, assure that the remote terminal will eventually become commonplace. Since the main use of the remote terminal is to retrieve and manipulate data, those who manage the agency's records and other information resources can expect an increasing demand on the part of management to computerize the agency's important data bases, particularly those that are dynamic in nature.

Rather than describing computers in accordance with their size, type, or operating characteristics, this chapter describes them in terms of the ways they are most often used for information storage and retrieval.

Computer index searching systems are those used to search index files where the indexing itself is performed manually. Indexers, using a guide such as a thesaurus of indexing terms, assign the indexing terms to the individual documents. The indexing terms are then usually coded, that is, converted to a numerical representation, and along with other pertinent data recorded in machine

A CRT TERMINAL WITH A LIGHT PEN

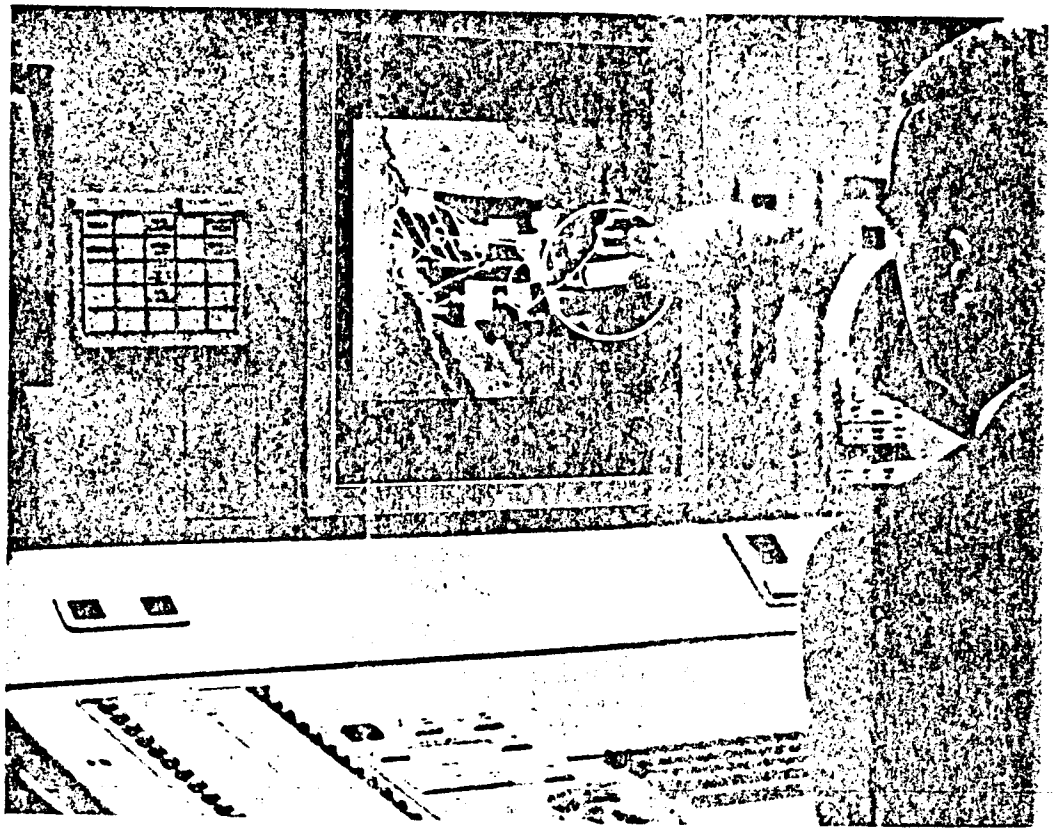


Figure 31

a well designed and properly operated computer indexing and searching system can be expected to perform about as well as those information retrieval systems where the indexing is done manually.

The major limitations of automatic indexing, searching, and preparing abstracts or extracts are the cost and the high degree of expertise required to design and operate such systems. However, the cost factor will become less critical as more and more offices move toward integrated information processing and retrieval systems that ultimately may include such features as computer-assisted document preparation and revision, computerized editing and preparation of the table of contents and index, and computerized printing. A copy of the same computer magnetic tape that goes to the Government Printing Office for use in automatic photocomposition and printing or is used to produce microform copy by COM equipment will also serve as input to the automatic indexing system, thereby eliminating one major cost—that of conversion of the information retrieval system input to machine language. These integrated information processing systems have one advantage that for many organizations may be far more important than the possible savings in cost—namely, the reduction in the period that elapses between the time an important event occurs, a fact is discovered, or a decision rendered, and the time the information is in the hands of those for whom it is destined or who may be searching for it.

Those persons interested in learning more on the subject should read NBS Monograph 91, Automatic Indexing: A State-of-the-Art Report, reissued February 1970 by the National Bureau of Standards (NBS) U.S. Department of Commerce.

Selective dissemination of information (SDI) systems are those that employ the computer or punched cards to provide individual users or user groups with tailormade announcements of new documents in their individual spheres of interest. The user's interest profile may be developed by having him look over the thesaurus of indexing terms and select those terms that reflect his areas of interest. The results are then recorded on a magnetic tape. Each time a new document is indexed, the indexing terms assigned the document or appearing in the abstract are compared with those stored on the user profile magnetic tape. In those instances where the requirements for a match are satisfied, the user is sent an announcement of the document, including its abstract, if any. Figure 33 illustrates an article announcement (abstract) card and a card used by the recipient to respond to the SDI system operators. Note that there are blocks on the recipient's response forms for him to use in indicating whether or not he wants to see the document and if not, why not, thereby providing the system operators with the necessary feedback.

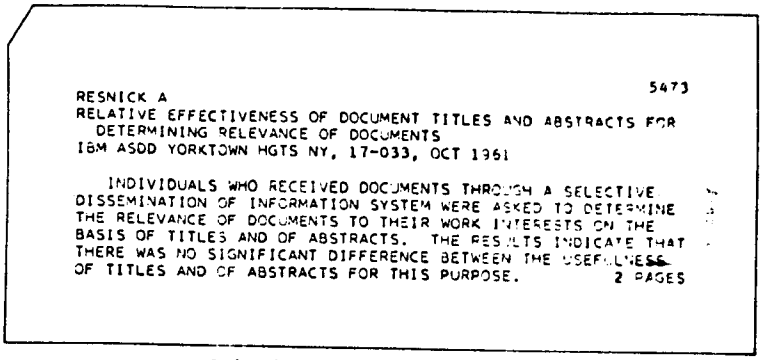
An interesting variation of the SDI technique is to develop interest profiles for major projects or programs, instead of for people, and to use the computer to keep the project director informed of any new documents on the subject.

While the costs for SDI systems are appreciable, the costs may not be considered unreasonable from management's point of view, particularly in the areas of scientific and technical research and development. However, scientists and engineers are not the only professionals having problems in wading through the tremendous volume of new documents made available to them, while at the same time trying to make sure they have not missed any documents that could have a major impact on their work.

The trend toward using group interest profiles rather than the profiles of individual users is resulting in less expensive and many times more practical SDI systems. SDI systems are especially valuable in providing the user with "peripheral vision" of information of direct interest to him, but which might be overlooked without the benefit of an SDI service.

Computer data storage and retrieval systems, sometimes referred to as data banks, are those used to store, retrieve, and manipulate large volumes of data (facts, numbers, letters, and symbols representing basic elements of information that can be processed or produced). Data bases may be either of two types or perhaps a mixture of the two: (1) recurrent or dynamic data, which is subject to change, and (2) noncurrent or static (archival) data relating to a unique event or representing an unchanging situation. The data base

SAMPLE SDI SYSTEM NOTIFICATION CARDS

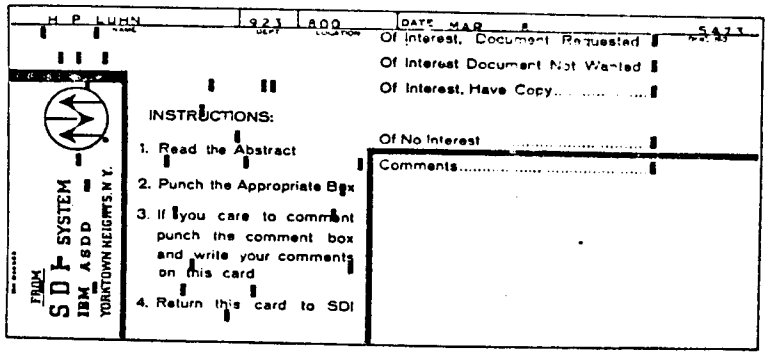


5473

RESNICK A
RELATIVE EFFECTIVENESS OF DOCUMENT TITLES AND ABSTRACTS FOR DETERMINING RELEVANCE OF DOCUMENTS
IBM ASDD YORKTOWN HGTS NY, 17-033, OCT 1961

INDIVIDUALS WHO RECEIVED DOCUMENTS THROUGH A SELECTIVE DISSEMINATION OF INFORMATION SYSTEM WERE ASKED TO DETERMINE THE RELEVANCE OF DOCUMENTS TO THEIR WORK INTERESTS ON THE BASIS OF TITLES AND OF ABSTRACTS. THE RESULTS INDICATE THAT THERE WAS NO SIGNIFICANT DIFFERENCE BETWEEN THE USEFULNESS OF TITLES AND OF ABSTRACTS FOR THIS PURPOSE. 2 PAGES

Article (Abstract) Announcement Card



H P LUHN | 923 | A00 | DATE MAR 8 | 5473

FROM SDI SYSTEM IBM ASDD YORKTOWN HEIGHTS N Y.

INSTRUCTIONS:
1. Read the Abstract
2. Punch the Appropriate Box
3. If you care to comment punch the comment box and write your comments on this card
4. Return this card to SDI

Of Interest, Document Requested
Of Interest Document Not Wanted
Of Interest, Have Copy
Of No Interest
Comments

Recipient's Response Card

Figure 33

may be specially created for information retrieval purposes, as in the case of weather data, or it may be used to serve multiple purposes. For example, census data is used for developing statistics and preparing reports as well as for information retrieval.

The social security and Federal income tax data bases are used mainly for automatic data processing purposes and only secondarily for information retrieval. Computerized management information systems also serve two purposes—to automatically produce reports and other communications and for information retrieval. It is the exception rather than the rule that a data bank is created and used solely for information retrieval. However, unless careful attention is given to the information retrieval needs in the planning and design of these multipurpose computer systems, there may be serious limitations or problems when later attempts are made to use the system for retrieving information.

For example some of the earlier ADP systems, in attempting to keep the machine record as short as possible, omitted such important data as the names of the individuals whose records were being maintained in the computer. Others were designed in such a way that individual items of data could not be selectively retrieved because the data was merely printed out in long lines without column headings. Sometimes the data was expressed in coded form, making it necessary for the user to refer to a special table to interpret the printout. Another problem, which is particularly critical at this time, is the lack of standardization or compatibility in data elements, thus making it difficult and sometimes impossible to exchange, compare, or combine data maintained in separate systems but relating to the same people, places, or things.

Unlike computer index searching systems and computer automatic indexing and searching systems, computer data storage and retrieval sys-