Information Sciences

JOSEPH BECKER AND ROBERT M. HAYES:
Information Storage and Retrieval

CHARLES P. BOURNE:
Methods of Information Handling

HAROLD BORKO AND ROBERT F. SIMMONS:
Automated Language Processing In preparation

ROBERT M. HAYES:
Mathematics of Information System In preparation

Methods of Information Handling

CHARLES P. BOURNE, Stanford Research Institute, Menlo Park, California

John Wiley & Sons, Inc., New York . London . Sydney

Library of Congress Catalog Card Number: 63-20628
Printed in the United States of America

Information sciences series

Information is the essential ingredient in decision making. The need for improved information systems in recent years has been made critical by the steady growth in size and complexity of organizations and data.

This series is designed to include books that are concerned with various aspects of communicating, utilizing, and storing digital and graphic information. It will embrace a broad spectrum of topics, such as information system theory and design, man-machine relationships, language data processing, artificial intelligence, mechanization of library processes, non-numerical applications of digital computers, storage and retrieval, automatic publishing, command and control, information display, and so on.

Information science may someday be a profession in its own right. The aim of this series is to bring together the interdisciplinary core of knowledge that is apt to form its foundation. Through this consolidation, it is expected that the series will grow to become the focal point for professional education in this field.

To my wife, Liz, who furnished the coffee and encouragement; and to our young children, Glen and Holly, who did the crayoning and snipping of much of the early manuscript material.

Preface

This book is meant to be an aid and reference work for those people who are interested in the design of information systems. Such information systems are typified by the traditional libraries, but the same problems of information processing, storage, and retrieval are present in many government and industrial organizations in many places besides the library. This book provides an illustration of the tools, equipment, and methodology that might be applied to those problems. Wherever possible, the chapters are liberally sprinkled with cost estimates, practical words of caution, and references to supporting literature. The chapters follow a somewhat logical progression, starting with the basics of file organization and moving up to methods of coding, notation, and machine-language representation; then through various manual systems and into the more complex equipment such as tabulating card and computer systems, and finally microfilm and micro-image equipment. These chapter divisions are somewhat artificial, but they did provide a convenient way to package the book contents.

Because of the nature of some of the material in the book, such as the equipment descriptions, the reader should expect that some of this material will be out of date by the time that he reads it. Unfortunately it can never be any other way. In the same way, the book was not meant to be an encyclopedia. Consequently, there will be some errors of omission, but these are felt to be slight.

Many of the references cited in this book are to the technical report literature. Those document numbers with the PB- prefix (e.g., PB-97 535) are cataloged by and available from the U.S. Department of Commerce, Office of Technical Services (OTS). Those document numbers with the AD- prefix (e.g., AD-270 942) are cataloged by and available from OTS and the Defense Documentation Center (formerly known as ASTIA).

I must sincerely acknowledge the support and assistance which was provided to me by the Stanford Research Institute for the preparation of this book. Particular credit must go to Mr. Donald Ford, who co-authored the chapter on special coding topics. Finally, my sincere thanks go to Mrs. Pauline Atherton for preparing the index to this book as well as providing critical and very helpful suggestions and reviews of the manuscript at several stages in its evolution.

CHARLES P. BOURNE

Menlo Park, California
August 1963

Contents

The World's Scientific Literature
The Classic Library Problems
Commercial and Government File Problems
The Emergence of Special and Complex File Systems
Some Fundamental File Problems
What the Rest of the Book Is About

1

The nature of the problem

Information handling and processing problems are present all around us in a variety of forms and in various degrees of complexity. This book describes many such problems, and suggests a large variety of tools and techniques that may be useful in the solution of these and similar problems. These tools may not be widely understood since it is only during the last few years that our very complex and technically oriented society has created a situation in which more and more people and organizations have become concerned with handling information and fewer with handling materials. We are generating more information and more paperwork; we are working with files of information that didn't exist or were insignificantly small in 1900.[1] Furthermore, we demand easier access to previously recorded information.

The result is a storm of paper, and fortunes in money and manpower spent solely on processing, storing, and retrieving information. The magnitude of this paper storm can be seen from a few isolated examples: (1) the Federal Government produces 25 billion pieces of paper per year and has accumulated enough records to fill 7.5 Pentagons, at a total cost of about 4,000,000,000 dollars a year; (2) military engineering documentation (research reports, manuals, drawings, etc.) now costs about 2,000,000,000 dollars per year; (3) approximately 12,000,000,000 to 15,000,000,000 checks pass through the nation's banks each year, with each check being handled 5 to 7 times; (4) more than 6,000,000 engineering drawings are made yearly for the military services, and there are more than 50,000,000 drawings on file; (5) approximately 30,000 technical journals (with several thousand new ones per year) publish more than 2,000,000 articles per year throughout the world in over 50 languages; (6) approximately 75,000 new book titles are published annually throughout the world; (7) some magazine publishers have combined mailing lists of up to 40,000,000 subscribers; (8) Federal offices maintain files for more than 160,000,000 fingerprints, 165,000,000 social security accounts, and 93,000,000 tax returns.

[1] **For example, fingerprints, motor vehicle registrations, social security and income tax records, patents, and engineering drawings.**

There is a need for methods or techniques to improve all aspects of this documentation problem of commerce and technology. The size of the mass of documentation, as well as the increasing dependence on documents, makes it necessary to find ways to reduce or maintain documentation costs, to increase the speed of transmission of information, and to maintain or reduce the amount of manual and clerical effort involved. In addition, new techniques are sought that could increase the productivity of the labor force. In response to these demands, there has been a steady development of techniques for the mechanization or automation of some of the documentation and retrieval functions.

Before we discuss specific problems and their solutions, some definitions and distihctions should be established in order to illuminate the material that follows. These distinctions often govern the selection of equipment and procedures to be used.

First, let us distinguish between data processing and data storage and retrieval. Both may involve a very large file or library of information; and no distinction is possible on the basis of size. Data processing systems have as their main goal the manipulation, replacement, alteration, or addition to the various items in the file. Library charge-out systems, customer utility accounts, and current airline reservation records are data processing examples. The data storage and retrieval systems have a different main goal: to store the file items for later re-use rather than modification, and most often to maintain the file items unaltered.

Second, let us distinguish among reference retrieval, document retrieval, fact retrieval, and information retrieval. Reference retrieval systems are typified by the library card catalog or other indexes, which yield a complete reference to a document in response to a general search request. Many of the mechanized retrieval systems, particularly the computer systems, provide reference retrieval only. Document retrieval systems go one step further and provide a complete copy of the document instead of just a citation or reference. Fact retrieval systems yield specific information (e.g., boiling point, tensile strength, chemical compounds) in response to a search request. Information retrieval systems are truly the most complex systems, since they must deal with concepts and abstractions, and provide specific thoughts or concepts rather than single facts or references. A true information retrieval system, for example, should be able to provide answers (not references) to a question such as "What is the difference between a point-contact transistor and a junction transistor?" Good answers exist for this question, and a true information retrieval system should be able to provide them. However, only a few information retrieval systems have yet been developed that come close to achieving this degree of performance.[2] The term information retrieval has long since become firmly established in common usage as the generic term that includes reference, document, and fact retrieval. Since it seems impossible to dislodge this usage, the term information retrieval is used throughout the rest of this book even though in most cases it will mean document retrieval.

And now in order to establish a proper frame of reference for some of the more powerful tools described in the following chapters, let us explore some of these problems in more detail.

[2] Green, B. F., et al., "Baseball: An Automatic Question-Answerer," *Proceedings of the 1961 Western Joint Computer Conference*, pp. 219–224 (Institute of Radio Engineers, New York, 1961).

THE WORLD'S SCIENTIFIC LITERATURE

The magnitude of the problem of the storage and retrieval of scientific and technical literature is frightening. Although the specifics of the problem may change from time to time. The magnitude of the problem is unlikely to decrease. Figure 1-1 shows the current amount, by national and linguistic origins, and degree of coverage by the abstracting and indexing services, of the world's periodical technical literature.[3] Note that this refers only to periodical literature. The amount of non-periodical literature (such as government reports, industrial research reports, and conference proceedings) has never been determined.

Total Volume of Publication

One of the most current and probably the most accurate estimates of the total volume of publication is a recent Library of Congress estimate that approximately 30,000 scientific and technical journals are published throughout the world. Others have suggested various totals between 50,000 and 100,000. The Library of Congress estimate, however, is based on a continuing census of the world's technical periodicals.

Using 30,000 journals as a starting point, and estimating 30 to 70 articles per journal per year, we are faced with some 900,000 to 2,100,000 articles published throughout the world every year.

National and Linguistic Origins of the Total Volume

Some very crude estimates have been made of the national and linguistic origins of the total volume of literature. English is still the predominant language, comprising about one-half of the total production. However, there are indications that Russian is moving into the position of second most popular scientific language, coming abreast of the traditional French and German. The United States still produces the greatest volume of literature, followed by Germany, France, and the United King-

[3] Bourne, C. P., "The World's Technical Journal Literature: An Estimate of the Volume, Origin, Language, Field, Indexing, and Abstracting," *American Documentation*, Vol. 13, No. 2, pp. 159–168 (April 1962).

Gottschalk, C. F. and W. F. Desmond, "Worldwide Census of Scientific and Technical Serials," *American Documentation*, Vol. 14, No. 3, July 1963 provides a more recent estimate.

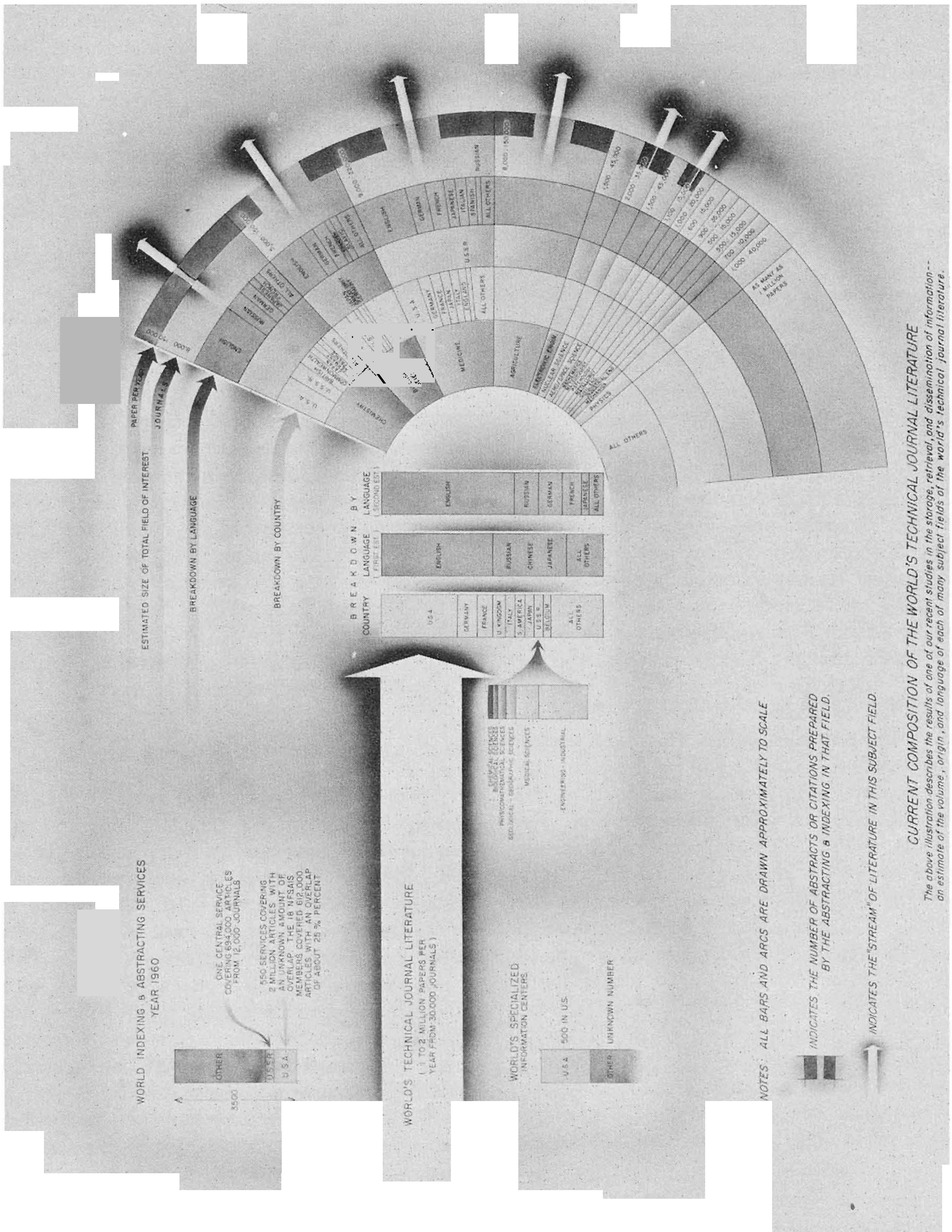


FIG. 1-1 Current composition of the world's technical journal literature.

dom in that order. However, the relative proportions for each country appear to differ markedly in the various specialty fields. The Soviet literature, for example, seems to be very prominent in chemistry, but relatively light in other fields such as medicine; this may be the true situation or it may reflect outdated information or errors in measurement. In any case, this pattern of national and linguistic origins should influence some of our academic and professional linguistic requirements, as well as our literature acquisition and abstracting policies.

The Magnitude and Characteristics of Several Subject Fields

The world's scientific community has recognized and defined various groupings of subjects or disciplines into "fields" such as chemistry, medicine, and mathematics. There is little that is definite about the boundaries of any of these fields, and in many cases there is overlap or complete inclusion of one subject field in another field. Nevertheless, a certain body of material or subject matter has been grouped together, usually around an academic tradition, a professional society, or an indexing and abstracting service, to form a single body of material. Artificial as the grouping may be, it exists; and the literature of each single group represents the normal span of interest or information resources for an individual working in that field. That is, a worker in any of these subject fields will see a certain volume of literature of interest to him; to him, this represents his technical information problem. The fact that his familiar reference tools cover a certain structured field of knowledge may also make it difficult for him to locate relevant information in other subject fields. It is also true that two workers in two different subject fields may also be looking at some of the same literature. To simplify the graphic portrayal, Fig. 1-1 shows distinct boundaries between neighboring subject fields. However, there was no intention of suggesting, for example, that there is a sharp boundary between chemistry and the biological sciences. In fact, the specialty field segments in Fig. 1-1 were arranged almost randomly.

Abstracting, Indexing, and Reference Services

There are currently about 3500 abstracting and indexing services throughout the world, with about 550 of them in the United States. There are also an estimated 3000 special information centers in the United States that maintain collections of information on special technical topics (e.g., Snow, Ice and Permafrost; Defense Metals; Air Pollution) for literature searching and reference services.[4] Most of these services or centers utilize conventional manual bibliographical tools; however, several of them are utilizing machine aids and other special techniques--most of which are described in this book.

As illustrated by the shaded arcs in Fig. 1-1, some specialty fields, such as chemistry, are well covered by abstracting and indexing services; other fields are badly covered. With the possible exception of chemistry, there is currently no practical mechanism, nor is there likely to be one in the next few years, for searching the entire world's literature in any specified subject field to answer a oearch request. Some indication of how far we are from the indexing coverage of the world's literature may be obtained by noting that in 1960, all of the major U.S. scientific abstracting and indexing services combined—services that constitute the membership of the National Federation of Science Abstracting and Indexing Services (NFSAIS)—covered only about 16,000 journals out of the world's estimated 30,000 scientific and technical journals.

THE CLASSIC LIBRARY PROBLEMS

The university libraries and some of the public libraries have grown to a very large size, and represent collections of several million books and periodicals. Figures 1-2 and 1-3 give an indication of the file size and accession rates for each of the U.S. public library systems and most of the U.S. college and university libraries, respectively.

Some efforts have been made to mechanize the literature searching in several major subject fields. The American Society for Metals, for example, has supported projects for several years for the machine searching of the metallurgical literature. The National Library of Medicine and the Chemical Abstracts Service are both planning to mechanize the searching of the literature of their subject fields.

[4] *Specialized Science Information Services in the United States. A Directory of Selected Specialized Information Services in the Physical and Biological Sciences,* Report NSF *61-68* of the National Science Foundation, Washington, D.C. (November 1961).

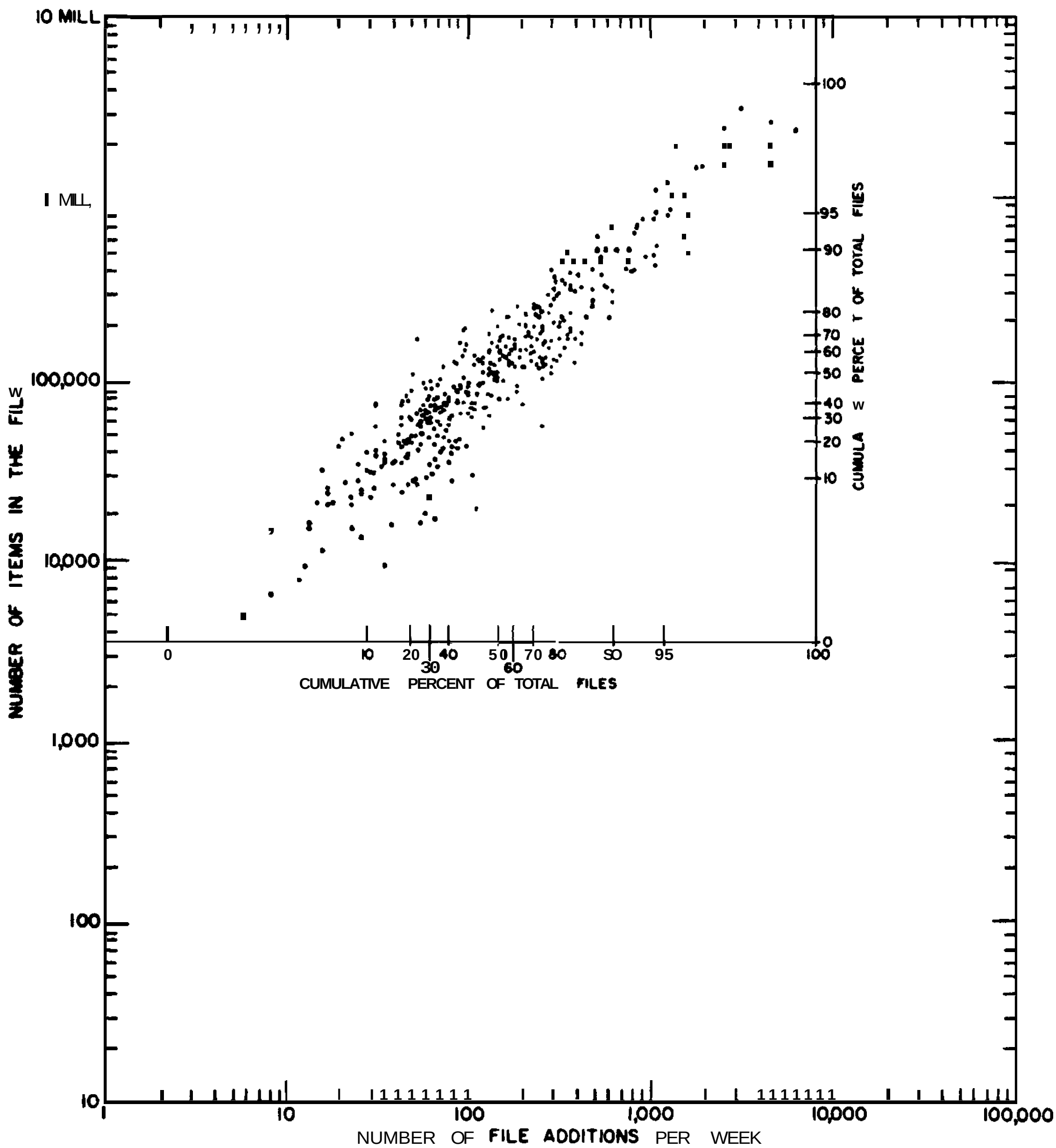


FIG. 1-2 U.S. public library systems—file size and accession rates. [Sources: Statirtics of Public LIbrary Systems in Cities with Populations of 100,000 or Mom: Fiscal Year 1958, U.S. Dept. of Health, Education ond Welfare, Office of Education, Circular 590 (June 1959); Statistics of Public Library Systems in Cities with Populations of 50,000 to 99,000: Fiscal Year 1958, U.S.. Dept. of Health, Education and Welfare, Office of Education, Circular 594 (July 1959); Public Library Statistics: 1944–45 (for cities with populations of 25,000 to 49,999). Federal Security Agency, Office of Education (1947).]

Projects are also under way for searching the legal literature, and the literature of several other subject fields. The magnitude of this type of searching task is reflected in the total amount of literature published annually in each field (see Fig. 1-1) and in the warehouse of abstracts or citations that each of the major abstracting and indexing services has accumulated during its history. Figure 1-4 shows the total number of abstracts or citations produced by each of several of the world's technical literature abstracting and indexing services, annually and throughout their publication history.

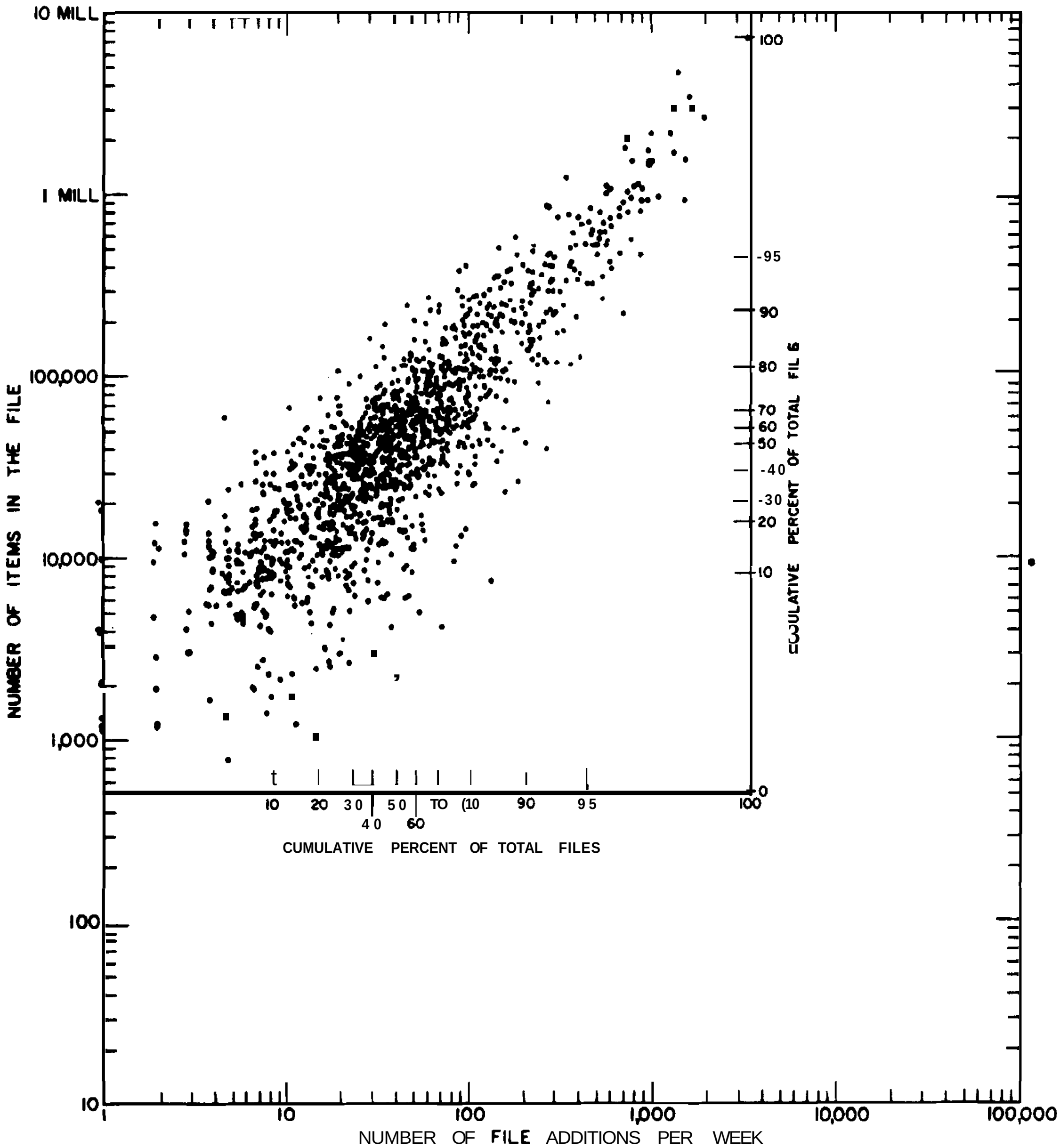


FIG. 1-3 U.S. college and university libraries—file size and accession rotes. [Source: ***Library Statistics*** of Colleges and Universities, ***1959–60;*** Part **I**: ***Institutional*** Data, U.S. Dept. of Health, Education and Welfare, Office of Education, J. C. Rather and D. C. Holladay, Report OE-15023 **(1961)**.]

COMMERCIAL AND GOVERNMENT FILE PROBLEMS

Libraries have no monopoly on storage and retrieval problems. Many commercial and government file problems are every bit as complex as the conventional library problems, and often involve more severe time and cost pressures. Here are some examples of these types of problems.

There are dozens of commercial firms in this country that make a business of maintaining and selling lists of names or organizations for direct mail promotion and advertising.5 Some of these firms have lists of **30** to 40 million names or organizations, often with each name extensively indexed to show the size of the town, the in-

[5] McComb, W. "Medical Mailing," *Data Processing,* Vol. 3, No. 1, pp. **15–17** (January **1961**).

dividual's credit rating, the type of automobile owned, and so on. For each client, the lists are made up from the master list in a manner tailored to suit the particular promotional needs, and may be as specific as lists of pediatricians in private practice in towns of 7000 population or more, or as general as the innocuous "occupant" type of list.[6] The mailing-list houses would like to be able to efficiently and inexpensively index their files to facilitate the generation of custom made lists which are more useful to their clients. A few of these firms can process more than 1,000,000 pieces of mail per day.

[6] The "occupant" type of file entry has the strong advantage in this case of normally not requiring any file maintenance unless the building at the occupant's address is destroyed and not replaced.

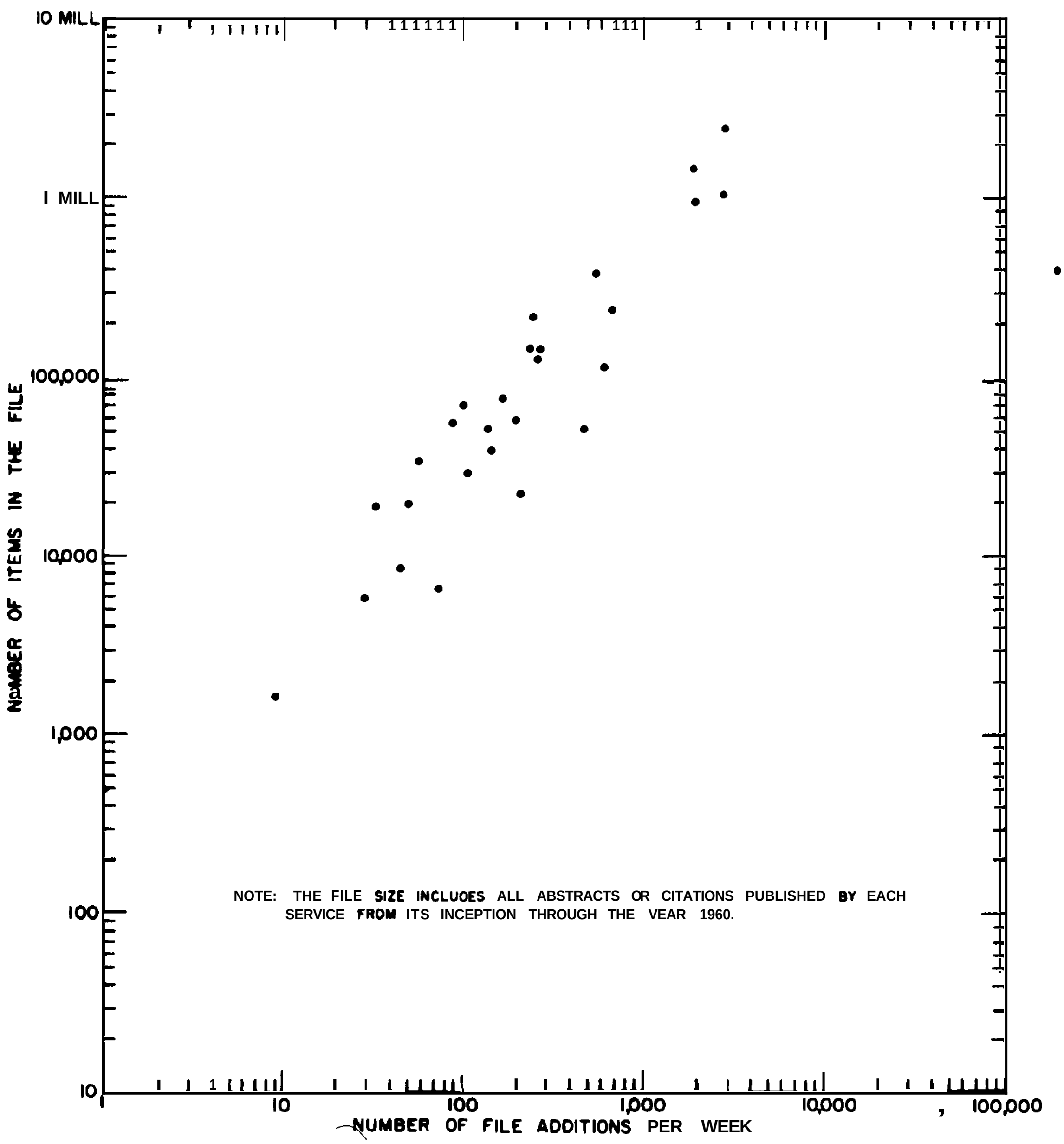


FIG. 1-4 Accumulated file sizes and current accession rates of the publications of wveral abstracting and indexing services. Note: The file size includes all abstracts or citations published by each service from its inception through the year 1960.

Each state, and several federal agencies maintain extensive collections of fingerprints, indexed in enough detail so that searches can be made to find the prints on file which match a print of unknown origin. Some states have as many as 5,000,000 prints on file, while the FBI file contains approximately 160,000,000 prints and is the largest.[7]

Nearly every firm that builds equipment or devices makes a complete series of engineering drawings for all aspects of that equipment, so that an identical equipment can be built at a later time, perhaps by a diierent organization. This is a very large task for complex projects or equipment. A complete description of a modem military aircraft, for example, may require approximately 80,000 different drawings. Some sort of a filing system must be established so that drawings for selected portions of the system can later be inspected as required. An aircraft manufacturer who develops and maintains many different pieces of complex equipment has file problems. So does the military service or airline that buys the equipment. The Navy Bureau of Aeronautics depository, for example, currently has about 5,000,000 engineering drawings on file for the equipment it has purchased or had developed in recent years.

A hospital patient's medical record may consist of handwritten notes, typewritten reports, inked graphs, strip-chart recordings, and correspondence. Hospitals generally keep the records of all of their patients, and an attempt is also made in all hospitals to catalog every patient's record according to the disease or type of injury as well as the part of the body that was affected. College courses and national standards examinations are required of the medical records librarians who do this type of indexing in each of the more than 5000 hospitals in the United States today. A city or county hospital may have files for approximately 250,000 patients, while the largest hospital complex, that of the U.S. Veterans Administration, has a file of medical records for approximately 8,000,000 ex-servicemen.

There are currently more than 1400 life insurance companies and 350 fire and casualty insurance companies in the United States, each of which maintains a file for each of its policies. There are, for example, over 250,000,000 life insurance policies in effect today.

There are more than 2500 title insurance companies in the United States whose business is to search all files of public legal documents to find all the mortgages, liens, and other legal actions pertinent to a client's real estate. The searches must be complete and accurate because the title company usually writes an insurance policy that essentially guarantees the client that they have provided complete documentation of the property's legal history.

There are approximately 20,000 municipalities in the United States that support a police force. A good record-keeping and file system is an essential tool in law enforcement. A police force may have files of registered fire arms, of crime reports, of photographs (mug shots), of fingerprints, of stolen property descriptions, of modus operandi reports, and of dossiers on known criminals. Extensive indexing and machine searching of the description and modus operandi of people with prior arrests has been done by some local and state agencies to assist in the identification of possible suspects.

THE EMERGENCE OF SPECIAL AND COMPLEX FILE SYSTEMS

In addition to the classic library problems and the commercial or government office problems, many new file problems arise from the technology of our modern age that do not fit into these categories. Some examples of these are given below.

The Defense Atomic Support Agency has been assigned the responsibility of creating within the Department of Defense a Damage Assessment Center whose primary purpose will be the assessment of United States, allied, and enemy forces and resources and their changes in the event of hostilities. Using a very large computer searching system, the Center must maintain a very large and comprehensive file on everything that could be considered a military resource throughout the entire world. The system utilizes a powerful information retrieval system to rapidly provide specific information in response to the inquiries of the Center's personnel.

The weather study satellites TIROS and NIMBUS are equipped with camera and television units to record and transmit pictures of the earth's cloud cover for weather forecasting on a global basis. TIROS pictures have been very useful for immediate weather forecasting, and the U.S. Weather Bureau is planning continuous weather satellite coverage of the world. Copies of these pictures are indexed and sent to a central file to be available for further research by the nation's meteorologists. Since the file items are generated at a rate of at least 18,000 pictures per satellite per month, it won't take long for the file to grow to a very large size. The research files at the National Weather Records Cen-

[7] Anon., "The FBI's Office Story," *The Office,* March 1960, p. 79.

ter already contain weather observations punched on 350,000,000 cards.

The Photographic Records and Services Division (PRSD) of the Department of Defense is the depository for most of the aerial photography acquired by the Department of Defense, and represents another film library problem. The PRSD library contains approximately 16,000,000 prints and original film indexed on a geographical basis, has an accession rate of about 1,000,000 pictures per year, and furnishes 2500 to 7500 prints per day.[8] Photographs from satellites are expected to increase the burden of this library even further.

The point of all of this should be clear by now. Storage and retrieval problems occur in many places besides the conventional libraries, and may be every bit as large and complex as the classic library problems.

SOME FUNDAMENTAL FILE PROBLEMS

Certain fundamental problems occur in varying degree in handling nearly every file, and especially in handling the large files, regardless of the type of material in the file or the method of file organization. Several of the major problems are discussed in some detail in the following sections. In these discussions, and throughout the rest of the book, a "file item" is defined as a smallest module or package of information to be handled as a unit. Books, patents, documents, insurance policies, medical records, and individual aerial photographs are examples of individual file items. A "file," as the word is used in this book, is defined as a homogeneous collection of a single type of file item.

Variances in Indexing

Few cataloging or indexing techniques avoid the possibility that the *same* document or file item may be indexed differently by different indexers—in fact, there is often a strong possibility that the same indexer would classify the same document differently on different days. This problem seems to increase with the number of indexers. It reflects such factors as: (1) differences in background and training of indexers, resulting in different appraisals of the subject material; (2) the amount of slack, variance, or ambiguity within the indexing system used; and (3) the relative subject importance of the material to be indexed; and (4) the variation and special language or jargon used in that material. Also, as the indexer becomes increasingly familiar with the special terminology and novel aspects of a body of material he changes his indexing habits accordingly.

One method for reducing the indexing variance is to take some of the slack out of the indexing system. The use of an authority list (i.e., glossary, dictionary, thesaurus) of subject headings helps to reduce some of this slack by providing written descriptions and clarifications of the various subject headings or terms used. A second method for reducing the indexing variance is to have each file item indexed by a number of different persons, each with a different background and subject specialty. An item so classified, on the basis of the composite indexing information, reflects the knowledge and thought patterns of a variety of file users. One large organization has used this technique very effectively for the indexing of technical documents by routing each document to 15 or 20 indexers, each representing a different scientific discipline and background.

At the output side of the system also, two people are likely to search the file in a different manner even though they are given identical search requirements. The reasons for this are essentially the same as the reasons for the variances in input indexing. Some relief can be obtained by providing indexing standards and practical on-the-job training in preferred search techniques.

Indexing Discontinuities Caused by Personnel Changes

The organization of a particular file cannot help but be influenced by the background and interests of the individual or group who established it. Because of this idiosyncrasy, the file often loses some of its effectiveness when the originators are replaced. New operators may find the file lacking in character and deficient in organization. They may also tend to index in a manner different from that of their predecessors. If differences are made unintentionally, it is probably because of the absence of a good system description and indexing procedures for the new operators to follow. For large files, adequate documentation of the rules and pro-

[8] *Systems Study of Photographic Records and Services Division, Document Data Processing Control, AN/GSQ-11 (V) Interim Report,* Eastman Kodak Co., Contract AF 30(602)-1698 (May 1958–January 1959), AD-225 499.

cedures to be followed should be provided, as well as close supervision and on-the-job training to acquaint novices with the system. For the majority of small files, it is probably easier to let the newcomer change the file organization into one that is more "comfortable" for him.

Loss of Material

Material gets lost or misplaced in any organized file, regardless of its size. As the file gets larger, however, it becomes more and more difficult to recover the lost items. In such applications as legal records and vital statistics, it is often necessary to maintain absolute integrity of the file; in such applications as advertising mailing lists, integrity is not as important.

Large libraries thread the catalog cards with a rod to reduce the possibility that cards will be permanently removed from the file, or taken out and mis-filed. This system is effective, but it is not foolproof and it cannot guarantee file integrity. In many cases, it is almost impossible to check to see if the file is actually intact. Systems which have a separate card or form for each file item (unit-record systems) are generally more susceptible to degradation of file integrity than are systems having less separable file media such as roll microfilm or bound volumes.

The use of color-coding or striping, may help to identify cards that are filed out of order; and often simplifies and speeds the re-filing operation. The use of special tabs on the tops of cards or folders may also help to show mis-filed items.

Difficulty of Modifying Original File Structures

Any file will suffer growing pains during its development, and will require revision or alteration of its structure. The task of revising a file is usually both intellectually and technically difficult. The reviser must decide how to change the file structure to correct for initial oversights, how to allow for changes in subject emphasis, development of new subject fields and terminology, and semantic shifts, and how to allow for future requirements that cannot be known beforehand. He must then decide how to change the markings on the books or documents, how to shift the location of the document if necessary, how to change the catalogs and indexes, and how to do all these things without interrupting the service and utility of the file.

If the file is attended by only one person (e.g., the desk file of an individual), revision may involve no more than re-labeling some manila envelopes and shuffling some papers in the desk drawers. If the organization is large, and the file is very complex, it may be necessary for a specially trained individual or staff to devote full attention to this problem, first accumulating statistics and information as a basis for the continued improvement of the file structure.

Loss of Familiarity with the File Contents

There is some chance that an indexer may lose his former familiarity with the contents and structure of a particular file, especially if he does not use the file. This problem may be reduced if the indexer physically works with the file and actively uses it for some tasks. This arrangement also improves the novice indexer's awareness of the file contents and could serve as an indoctrination or as on-the-job training for new indexers. The indexers can be introduced to the file in many ways, the most common way being for the indexer to conduct some of the routine file searching operations. Another way is to require the indexer to periodically perform historical or state-of-the-art research that will require the extensive use of the file.

Loss of Familiarity with the Subject Matter

The quality of the indexing of files of particularly complex subjects such as chemistry, electronics, or nuclear physics is usually governed by the indexer's knowledge of the subject. Successful indexers and information specialists in the technical information fields usually have their basic training in the subject field that they are indexing, with additional training in indexing and searching techniques. Furthermore, better results are obtained if the indexer is active in the subject field, rather than being an observer. The active worker is usually better informed about new and current technology, and is aware of trends in development. In addition, he appreciates from experience the' needs of the file user. The full-time indexer or information specialist tends to fall behind in his understanding of the technology of his subject. Motivations, ambitions, and communication channels are partly responsible for this.[9]

A variety of ways exist by which organizations can develop or obtain indexers who also work in special subject fields. Three useful methods are: (1) rotate personnel between technical work and information services work (e.g., alternate between the two types of work at 4-month intervals); (2) use the regular technical staff for indexing on a part-time basis (e.g., 4 hours of indexing or abstracting per week per man); (3) decentralize the management of the file to enable working technical specialists to organize and maintain those parts of the file corresponding to their special fields of interest. In this last case, each technical specialist would still devote the greater part of his effort to technical endeavors and responsibilities, but he would provide the guidance and leadership for the file clerk charged with maintaining that segment of the file.

File Purging

A program of systematic review must be adopted in nearly all situations to prevent a file from becoming cluttered with useless or outdated records that hinder efficient use of the file and waste space and equipment. In practice, the purging program often consists of infrequent massive assaults, in which the entire file is examined in one pass and a mass of material is discharged for destruction. For a large organization, this may result in tons of paper being removed from the files. However, in place of such massive and dramatic operations, it is much more reasonable to attack the problem at the source, by carefully screening the information to be filed, rejecting some material and marking other material for later review, transfer to less expensive storage facilities, or automatic destruction after a specified time interval. Purging may be done at regular intervals, or it may be done continuously and automatically as part of the system design.[10] Standards for records retention, based usually on statute-of-limitation considerations, have been established and applied for many fields of business.[11] Some large libraries have inaugurated book retirement programs. However, the problem of records retention for technical information is complicated by the variable and often unknown content and value of each particular file item. Efforts have been made to purge technical files on the basis of the age of the file item, but the useful life of a particular technical document can usually not be predicted to the satisfaction of the file users. There have been efforts to describe the useful "half-life" of a technical paper in various subject fields. In any case, files of any type or size must usually undergo some systematic weeding or thinning in order to maintain their effectiveness for the user.

Auxiliary or Snag Files

Large files can seldom grow without developing some small auxiliary or snag files to represent the items that will not fit in the organization of the main file. How many file systems can get along without the category "miscellaneous"? Just as splinter political parties develop for a special interest, and then disappear, or become absorbed by one or more of the main parties, so do the splinter files often become absorbed by the main files. They may exist because of certain weaknesses or deficiencies in the main file organization, and then become absorbed or replaced as the deficiencies in the main file are removed. Snag files may also develop when the file system must perform a function not in the original design. In these cases, the special-purpose snag files have a strong reason for existence, and will probably not be absorbed by the main file.

WHAT THE REST OF THE BOOK IS ABOUT

Now that the general information retrieval problem has been introduced, the rest of the book will be devoted to describing and suggesting a variety of tools, techniques, and hardware to help the reader

[9] Most organizations, for example, make a practice of sending representatives of their active technical staff to attend national conferences and other meetings, but are very reluctant to send staff personnel such as the indexers or information officers to the same or similar meeting.

[10] For example, piles of papers and references on a worker's desk may often be self-purging by a process of sedimentary deposition. That is, after a period of time, the papers which are used most often and are of active interest to the worker, end up on the top of the pile. The less useful papers usually end up on the bottom. The file has thus purged itself.

[11] *Retention and Preservation of Records with Destruction Schedules*, 6th ed. (Record Controls, Inc., Chicago, Illinois, 1961).

Wheelan, R. B., *Corporate Records Retention, Vol. 1: A Guide to U.S. Federal Requirements* (Controllership Foundation, Inc., New York, 1958).

with the solution of his specific problems. The division of the remaining material of this book into specific chapters may appear to be rather arbitrary —it is. However, the chosen approach of dividing the material into chapters corresponding to specific types of hardware (e.g., microfilm, computers, and punched card systems) seemed to be a useful compromise. All of the text material is extensively indexed, and a great many references are provided for most of the topics discussed. Hopefully, this will provide the reader with an easy route to the pertinent material in this book or in the open literature.

Fundamental Methods of Indexing Subject Content
Methods of Subject Index Display
Other Types of Indexes
Criteria for the Selection of an Indexing System

— 2

Classification and indexing: The organization of information

One of the first considerations in the design of a large file system is the way in which the information is to be organized and assembled. The methods of file organization are extremely important to the proper functioning and performance of any retrieval system—regardless of whether it is mechanized. The highly mechanized file systems need at least as good a file organization as the manual systems. Otherwise, the machine system may have the dubious distinction of making mistakes 1000 times faster than the manual system.

Mechanized systems can operate with almost any type of file arrangement. The selection of a particular file arrangement does not necessarily force the selection of a specific type of equipment, and vice versa.

After choosing the method of file organization, a secondary question is how the individual file items are to be coded for subsequent storage and processing by manual or machine methods. The organization problem is discussed in this chapter, and the coding problems (i.e., the symbolic representations) are discussed in Chapt. 3.

A single chapter cannot describe all of the different methods that are available for organizing collections of information on the basis of their subject content; nor can it give much practical advice on the use of these methods. However, it can describe the fundamental methods, the differences between these fundamental methods, and which of these methods have been mechanized. Some comments can also be made to describe some of the special search tools and display techniques used with these methods. Since the methods of indexing documents by the bibliographical data (e.g., author, date, publisher, etc.) are relatively straightforward and well described by the conventional library rules for descriptive cataloging, they will not be discussed further here. The attention of this chapter will be directed solely to subject analysis, that is, the organization of the file in accordance with the subject material in the file items.

There did not seem to be any simple way to satisfactorily divide this chapter into separate sections, each of which handled a fundamentally different type of indexing technique. No real fundamental differences could be found among the many techniques examined. There seemed to be a continuum or merging of one method into another, rather than distinct differences. In many cases, it was extremely difficult to determine the precise difference between subject headings, descriptors, *Uniterms,* and all the other indicators of subject content. Consequently, an alternative approach was taken—that of arranging the methods by increasing degree of control exercised on the growth and use of the indexing language. This approach was suggested earlier by Vickery, and has been somewhat modified in this discussion.[1] Conveniently, this also seems to correspond to the order of increasing intellectual effort expended by the indexer, and the increasing difficulty of mechanization. Some of the many possible steps in this continuum of indexing complexity are described in more detail in

[1] Vickery, *B. C., On Retrieval System Theory, Chapt. 4* (Butterworth & Co., London, 1961).

the following section; hopefully in such a way that their indexing strength can be studied independently of their particular notation, form of display, or special jargon. The rest of the chapter should illustrate the point that the basic aspects of subject indexing, notation, and display can be treated almost separately from each other. That is, any basic method of subject indexing can be used in conjunction with almost any method of coding, notation, or display. Further comments on the mechanization of the indexing operations are given in Chapts. 6 and 7.

FUNDAMENTAL METHODS OF INDEXING SUBJECT CONTENT

There is no precise way to show that a method of indexing becomes more powerful as an increasing amount of control is exercised over the indexing language, but this seems intuitively to be true. Various indexing techniques will be described in this section in what is felt to be an approximate order of increasing indexing power or effectiveness for retrieval purposes.

Words Chosen from Title or Text but Common Words Omitted

The first and simplest indexing technique is to lift or extract some words from the title of the document being examined and use these terms as indicators of the subject content of the document. But, since the title may not prove to be an accurate indicator of the document contents, a better approach would be to extract words from the abstract of that document—and, better yet-extract the words from the entire text or body of the document. For the purposes of the discussion of subject indexing in this section, we shall assume that the entire text of the document is being examined.

In this most elementary form of subject indexing, an attempt is made to indicate the contents of the document by a list of words extracted from the text. The only control on the indexing is to provide a list of common words, such as the ones given below, that are to be avoided since they have little value for indexing.

of	more	special
be	going	several
if	taken	another
up	often	produced
to	study	possible
on	great	approach
an	toward	problems
it	during	something
the	around	associated
and	before	variations
our	system	conditions
for	better	development
few	higher	demonstration
with	others	determination
some	because	characteristic

The search tool is usually an alphabetical list of the terms used to date, with an indication of what documents are associated with each of the terms. This list may be displayed in many forms, ranging from printed lists to separate record cards for each term. The document may be represented by many means, such as code numbers (accession numbers) posted to the term cards, or an abbreviated citation in the printed list.

The increasing popularity of such an elementary indexing procedure seems to be due in large measure to the fact that each document can be indexed rather quickly by people who do not have a complete grasp of the subject material which would be needed to classify it more exactly. At first glance, it may appear to be a very simple and useful approach—and in some cases it is. But, problems soon arise in the control of the indexing terminology. Synonyms, plural forms, generically related terms, and other points begin to cause difficulties in indexing, and searching, and the user then finds himself switching to another type of system or building in more controls such as those mentioned later in this chapter.

One of the earlier formalizations of an indexing system based on words extracted from the text was the Uniterm Coordinate Indexing System, proposed in 1952.[2] These early Uniterm ("unit-terms") systems did not stress vocabulary control and even suggested that the indexer was not required to create and maintain a list of approved Uniterms.[3] It was also suggested that the indexer would rarely have to consult the index (search tool) as he assigned terms to the new material.

The Uniterm concept was initially developed and

[2] Taube, M., et al., *Studies in Coordinate Indexing,* Vol. 1 (Documentation, Inc., Washington, D.C., 1953).

[3] *Installation Manual for the Uniterm System of Coordinate Indexing,* Armed Services Technical Information Agency (ASTIA) (Arlington Hall, Virginia, October 1953).

operated as a manual system. Since 1953, there has been much interest in this approach, with scores of libraries establishing manual Uniterm systems.[4]

The basic word units in the Uniterm system usually consist of single-word concepts or ideas (e.g., methane, transistors, B-47) although multiple-word units are used (e.g., human engineering, *Indian* Ocean, surface-to-air missiles). Nouns, adjectives, equipment model numbers, and proper names are also used as indexing terms. In any case, the indexer writes out a list of words from the text that he thinks are representative of the contents of the document. The sequence or relationship of one Uniterm to another is not considered. Each Uniterm is considered to be of equal value so that no indexing terms are subordinated or related to any other terms.

The indexing terms may be assigned so liberally that one document has as many as 30 to 40 Uniterms. There is no standard dictionary of Uniterms for universal application to a number of different files—just as a standard Universal Decimal Classification exists in published form and can be applied to a number of different files. The dictionary for any new installation often starts with zero entries, and grows as items are indexed and different terms are added to the vocabulary.[5]

A practice of liberal or uncontrolled assignment of terms has raised difficulties for many users, and some indexers are now exercising more control on their descriptor vocabulary than they have in the past. In some cases, a degree of control may be achieved by deriving the Uniterm vocabulary from an existing subject-heading list. The present Armed Services Technical Information Agency (ASTIA) system, for example, employs about 7000 Uniterms of which at least 80 percent consist of terms formerly contained in a list of subject headings devised by its predecessor, the Technical Information Division in the Library of Congress.[6] The same list of ASTIA Uniterms also corresponds very closely to the 13,000 subject headings used by the Atomic Energy Commission.[7]

The stated operating procedures for the Uniterm system suggest that each document to be indexed is given a unique number (a serial or accession number), which is then used throughout the system to represent that particular document. Figure 2-1 illustrates this point. Some attempts have been made to mechanize the preparation of these term cards with punched card or computer equipment and this is discussed later, in Chapts. 6 and 7.

Searching is done by translating the search question into specific Uniterms that represent the substance of the search request, extracting the records

[4] Taube, M., et al., "Unit Terms in Coordinate Indexing," American Documentation, Vol. 3, No. 4, pp. 213-218 (October 1952).

Gull, C. D., "Alphabetic Subject Indexes and Coordinate Indexes: An Experimental Comparison," College and Research Libraries, Vol. 14, No. 3, pp. 276–281 (July 1953).

Morris, J., "Evolution or Involution? A Note Critical of the Uniterm System of Indexing," Journal of Cataloging and Classification, Vol. 10, pp. 111-118 (July 1954) (M. Taube's reply on p. 119).

Randall, G. E., "Practicality of Coordinate Indexing," College and Research Libraries, Vol. 15, No. 4, pp. 417-419 (October 1954).

Warheit, I., "A Study of Coordinate Indexing as Applied to U.S. AEC Reports," College and Research Libraries, Vol. 16, No. 3, pp. 278-284 (July 1955) (Taube's reply on p. 284).

Sanford, J. A., and F. R. Theriault, "Problems in the Application of Uniterm Coordinate Indexing," College and Research Libraries, Vol. 17, No. 1, pp. 19–23 (January 1956).

Francisco, R., "Use of the Uniterm Coordinate Indexing System in a Large Industrial Concern," Special Libraries, Vol. 47, No. 3, pp. 117-123 (March 1956).

Armed Services Technical Information Agency, Conference on Multiple Aspect Searching for Information Retrieval, Washington, D.C., *February* 12–13, 1967 (December 1957), AD-147 491.

Whaley, F. R., "A Deep Index for Internal Technical Reports," in J. H. Shera et al., editors, Information Systems in Documentation (Interscience Publishers, New York, 1957).

Wadington, J. P., "Unit Concept Coordinate Indexing," American Documentation, Vol. 9, No. 2, pp. 107-113 (April (1958).

Costello, J. C., Jr., and E. Wall, "Recent Improvements in Techniques for Storing and Retrieving Information," in Studies in Coordinate Indexing, Vol. 5 (Documentation, Inc., Washington, D.C., 1959).

Jaster, J. J., B. R. Murray, and M. Taube, The State of the Art of Coordinate Indexing, a report prepared by Documentation, Inc., Washington, D.C. (February 1982).

[5] Schultz, C. K., and C. A. Shepherd, "A Computer Analysis of the Merck, Sharp, & Dohme Research Labs Indexing System," American Documentation, Vol. 12, No. 2, pp. 83-92 (April 1961).

[6] Dubester, H. J., "Mechanization and Subject Headings," Library Resources and Technical Services, Vol. 6, No. 3, pp. 230–234 (Summer 1962).

During the time that this book was in press, the official name of ASTIA was changed to the Defense Documentation Center for Scientific and Technical Information (DDC). The earlier name is used throughout this book.

[7] Hammond, W., and S. Rosenborg, *Experimental* Study of Convertibility Between Large Technical Indexing Vocabularies, Technical Report IR-1 of the Datatrol Corp., Silver Spring, Maryland (August 1962).

(cards) for each of the specified Uniterms, and noting the document numbers listed on each card. To answer the question, '(What documents do we have in the file that discuss the human engineering of display systems?"—the searcher would look for the document numbers that are common to the two cards human engineering and display. Figure 2-1 shows two representative cards that would be examined during such a search. In this case, the document numbers 670, 1380, 3451, and others, would appear on both cards and may provide relevant answers to the search question. It is quite possible, however, that some of the indicated documents, even though they are represented by the prescribed terms, may not be a correct answer to the search question. For example, given the question, "What documents do we have in the file that discuss the propagation of round-off errors in machine computations?" a report on '(Indicated Errors in the Propagation Computations" would be selected by a search that used the Uniterms propagation, errors, and computations, even though it is probably irrelevant. This problem of false drops has been suggested as one of the main drawbacks of the basic Uniterm scheme, but it often turns out to be a relatively minor problem in practice. When this type of problem becomes acute, it can be ameliorated by the use of auxiliary indicators ("links" and "roles") described later in this chapter which show the relationships between terms.

Also included in this same class of indexing techniques are all the present methods of permuted title indexing called Key-Word-in-Context (KWIC) **indexing.**[8] This type of permutation indexing, whether performed manually[9] or on data processing equipment (see Chapts. 6 and 7) is a technique for displaying a sequence of indexing terms derived from a title or portion of a document to be indexed. Its distinguishing feature is its form of display. Because the index term is shown in context, its indexing power is slightly greater (but only slightly) than the simplest of indexing schemes that lift words from the text. It might be considered a first rough draft of a more comprehensive indexing or classification scheme.

[8] This type of indexing has been referred to as "Key-Word-in-Context (KWIC) Indexing" because the method of display of the printed index actually shows the key words in some context.

[9] Janaske, P. C., "Manual Preparation of a Permuted-Title Index" *American Institute of Biological Sciences Bulletin,* Vol. 12, No. 6, pp. 53-54, December 1962.

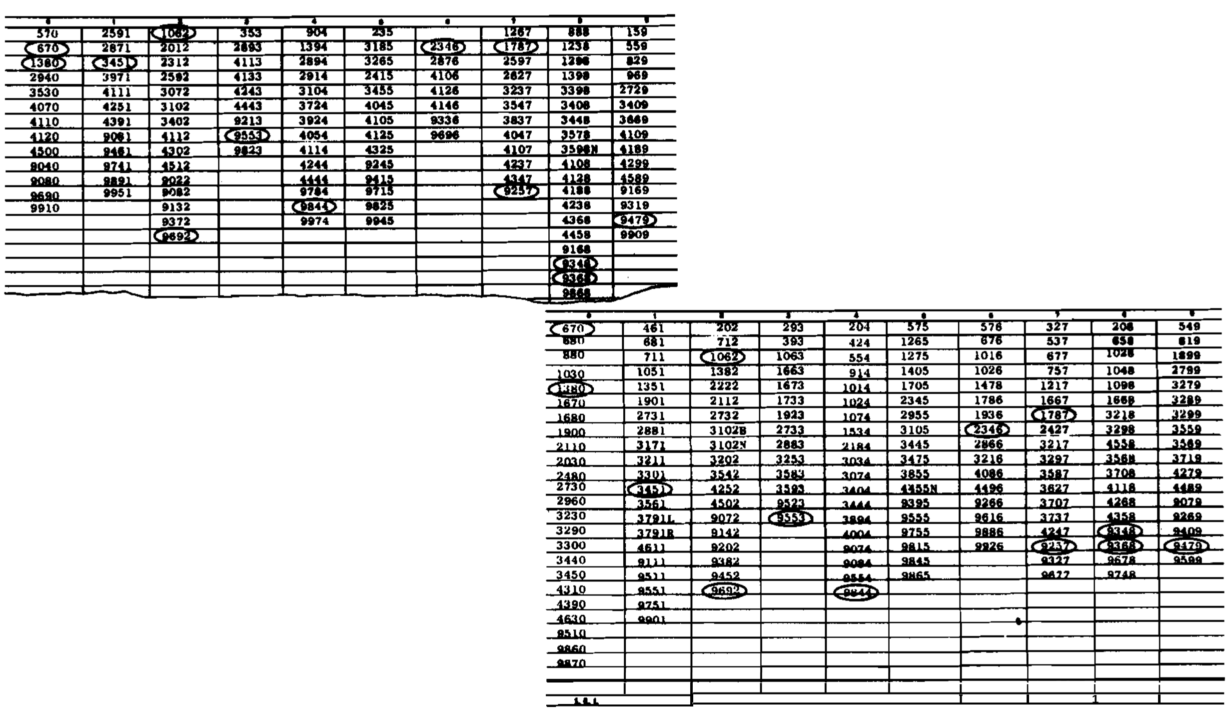


FIG. 2-1 Representative Uniterm terminaldigit term cards. Top: display; bottom: human engineering.

The early development and application of this method, and its relationship to other similar schemes is well described by the following footnote from a System Development Corporation report.[10] "Permutation indexing is a machine system developed by Herbert Ohlman of the System Development Corporation using fixed-field punched cards and extended to a field-free punched card system by Lewis Hart of the same organization. It is related very closely to the 'rotated file' developed at the Chemical-Biological Coordination Center, to the 'KWIC' system of H. P. Luhn at IBM, to the 'rotational indexing' developed by H. Snowden Marshall at Rocketdyne, and to the 'redundant self-indexed information recording' of Gene Bussey at Sandia Corporation. There are undoubtedly other similar developments—among the related manual systems are extractive filing as employed at Battelle Memorial Institute, 'correlative indexing' as propounded by C. L. Bernier of Chemical Abstracts, and bibliography compilation as developed by Major D. B. Netherwood at Wright Air Development Center." The first published reports to describe the implementation of this technique with data processing equipment came independently and simultaneously in 1958 from IBM and SDC (System Development Corporation).[11] A later publication reported that the Central Intelligence Agency had been using permuted title word indexing since 1953.[12] The methodology is continually being improved and is currently in use by many companies and publishers for indexes and title announcement bulletins.[13] This type of indexing is useful primarily for current awareness reporting (e.g., Chemical Titles, *Keywords* Index to U.S. Government Technical Reports, and BASIC, a semi-monthly title index to Biological *Abstracts*) and for the one-time indexing of relatively small and specialized collections of information where the titles of the items are particularly descriptive and meaningful (e.g., Dissertations in Physics). An example of such a display is given in the sample page of a KWIC index shown in Fig. 2-2. This particular illustration, from one of the earliest KWIC indexes, shows how the words of a single line of text are shifted or re-arranged horizontally, and then listed in alphabetic order by each of the significant words in that line.[14] In this particular figure, selected extracts from three sections of the index were chosen to show how the title "Punched Card and Machine" was listed in three different places in the index: under punched, card, and machine. The associated number, 5258, is the serial number of a full bibliographic description of the article printed in another section of the index. Portions of some of the longer lines of text have been shifted off the right end of the line and have looped around to be printed on the left side of the line.

The usefulness of this type of indexing depends in large measure on how well a document title or sentence describes the document's contents. The communication effectiveness of extracted document titles and sentences has not been clearly established yet.[15] It is becoming more frequent practice for

[10] Citron, J., L. Hart, and H. Ohlman, ***A Permutation Index to the Preprints of the International Conference on Scientific Information,*** a December 1959 revision of a November 1958 report by System Development Corporation, Santa Monica, California.

[11] ***Bibliography and Auto-Index, Literature on Information Retrieval and Machine Translation,*** an un-numbered report of the IBM Service Bureau Corporation, New York, 1st ed. September 1958, 2nd ed. June 1959.

SDC report cited previously in this section, fn. 10.

[12] Veilleux, M. P., "Permuted Title Word Indexing Procedures for a Man-Machine Syatem," in ***Machine Indexing: Progress and Problem,*** pp. 77–111 (American University, Washington, D.C., 1962).

[13] Luhn, H. P., "Key-Word-In-Context Index for Technical Literature (KWIC Index)," ***American Documentation,*** Vol. 11, No. 4, pp. 288–295 (October 1960).

Luhn, H. P., "The Automatic Derivation of Information Retrieval Encodements from Machine-Readable Texts," Chapt. 45 in ***Information Retrieval and Machine Translation,*** A. Kent, editor (Interscience Publishers, New York, 1960).

Kennedy, R. A., "Library Applications of Permutation Indexing," ***Journal of Chemical Documentation,*** Vol. 2, No. 3, pp. 181–185 (July 1962).

Kennedy, R. A., "Mechanized Title Word Indexing of Internal Reports," in ***Machine Indexing: Progress and Problems,*** pp. 112–132 (American University, Washington, D.C., 1962).

Turner, L. D., and J. H. Kennedy, ***System of Automatic Processing and Indexing of Reports,*** Report UCRL-6510 of the Lawrence Radiation Laboratory, Livermore, California (July 1961).

[14] SDC report cited earlier in this section, fn. 10.

[15] Resnick, A., "Relative Effectiveness of Document Titles and Abstracts for Determining Relevance of Documents," ***Science,*** Vol. 134, No. 3484, pp. 1004–1006 (October 6, 1961).

Maizell, R. E., "Value of Titles for Indexing Purposes," ***Revue de la Documentation,*** Vol. 27, No. 3, pp. 126127 (August 1960).

Lakhuti, D. G., and N. A. Stokolova, "The Problem of Searching Chemical Abstracts by Title," ***Foreign Develop ments in Machine Translation and Information Processing,***

•
•
•

```
     5421 FIGURE 2A TERM CODE IN QUESTION CARD /
ALYSIS WAS MADE BY PUNCHING A HOLLERITH CARD /                   1278 THE AN
           6046 TAYLOR - STENCIL ASPECT CARD /

           ——> 5258 PUNCHED CARD AND MACHINE •

   5419 THE COMAC , AN EFFICIENT PUNCHED CARD COLLATING SYSTEM FOR THE STORAGE -
                          5419 PUNCHED CARD COLLATION /

/                   4075 EDGE - PUNCHED CARD INADEQUATE IN LANGUAGE CAPABILITIES
                         1156 THE DIARY CARD INVESTIGATION *

     1169 APPENDIX 3 DISCUSSION OF DIARY CARD LAYOUT /

                  5091 MARGINAL - PUNCHED CARD PRODUCED ON IBM EQUIPMENT /

                                    5427 CARD REPRODUCTION AND PRINT OUT /
```

•
•
•

```
            ——> 5258 PUNCHED CARD AND MACHINE *
                                    2195 MACHINE - PUNCHED CARDS /
OORDINATED DOCUMENTATION USING HAND AND MACHINE -       2195 POSSIBILITIES OF C
                 4038 METHOD - SUCH AS MACHINE - SORTING TO COMPUTERS /
                                    5134 MACHINE APPLICATION +
                                    5135 MACHINE COMPARIOSN *
                              5340 SOME MACHINE DESIGN PROSPECTS /
                                    5280 MACHINE DISCLOSURES •
NG THE HARVARD -      5137 LINGUISTIC AND MACHINE METHODS FOR COMPILING AND UPDATI
                  5405 RAPID METHOD FOR MACHINE RETRIEVAL /
                5161 THE FEASIBILITY OF MACHINE SEARCHING OF ENGLISH TEXTS .
EARCH FILES /     5350 APPENDIX D USE OF MACHINES AS AIDS IN THE PREPARATION OF S
                                    5335 MACHINES WITH LEARNING ABILITY /
```

•
•
•

```
ONSIDERATIONS IN ESTABLISHING THE FIELD PUNCHED - CARD SYSTEM * 4087 PRACTICAL C
                                   4010 PUNCHED CARD SYSTEM /
                              ——> 5258 PUNCHED CARD AND MACHINE +
                                   5419 PUNCHED CARD COLLATION /
                                   5420 PUNCHED CARD SEARCHING +
                         5091 MARGINAL - PUNCHED CARD PRODUCED ON IBM EQUIPMENT /
ORAGE -     5419 THE COMAC , AN EFFICIENT PUNCHED CARD COLLATING SYSTEM FOR THE ST
             2090 SUBSCRIPTION SERVICE OF PUNCHED CARDS /
                             2195 MACHINE - PUNCHED CARDS /
          5316 FIGURE B5  OUTLINE OF THE PUNCHED CARD USED IN VS3 /
BILITIES /                   4075 EDGE - PUNCHED CARD INADEQUATE IN LANGUAGE CAPA

            1278 THE ANALYSIS WAS MADE BY PUNCHING A HOLLERITH CARD /
```

•
•
•

FIG. 2-2 Representative sample of a permutation or Key-Word-in-Context index.

indexers to add words or terms to otherwise barren titles in order to provide useful indexing points (sometimes called "title enrichment"). This leads the basic KWIC scheme into a more powerful method of indexing in which terms are added that are implied but not specifically stated in the title or text.

Figure **2-3** shows an example of a more recent form of a KWIC index, which uses a slightly different method of display.[16] In this case, the titles are left intact with no shifting along the line, and there is no restriction on the number of lines taken to describe the full title. The key words are copied off to the left margin of the title, with one entry assembled for each significant word. This makes the index somewhat easier to scan, and provides the full title instead of a one-line segment. The serial numbers again refer to full citations that are printed in another section of the index. This particular index listed the titles under an average of 6 key words per title. The key words selected were the remainder after a computer had rejected any of approximately **300** forbidden words that had been programed into the machine. The same index also includes an alphabetical listing of the reports by corporate author.

One operational disadvantage of KWIC indexing is the relatively bulky index that is developed for a given collection. Depending upon the amount of information provided, and the manner of presentation, the index may be printed to hold an over-all average of **10** to 25 titles per page.

Similar to the KWIC system, but providing much more information in context, is the "clue word" or "key word" extraction technique, in which key words are extracted from the title and text material to become the index terms. As with the KWIC method, only the specific subject words that appear in the document itself are used for indexing—and very few terms are assigned that do not occur in the document. This key word or extractive approach was put into use at about the same time that the first KWIC system was started and differs from KWIC primarily in its method and amount of display. Its display consists of a unit-record called an "extract," which is a condensation or abbreviation of the document (see Fig. 2-4). This extract is made by extracting or lifting key sections from the document to form a summary collection of statements.[17] Bibliographical data and other key subject words are also lifted to go with the summary. The indexer marks the words and sentences in the original document that summarize the document contents, and this marked material is transferred to form another unit record, such as a 5-by-8-inch card. These new cards are reproduced and filed under each of the key words that were marked (e.g., names, organization, subjects). In this way, a copy of an extract may appear under many headings in the file system. A technical literature file may require each original extract to be filed, on the average, in **10** to 15 different locations. For searching purposes, a person can go to the file section under the subject heading or clue word of interest, and immediately browse through a file of all relevant summary or extract cards. The cards contain enough text to be useful, and the subject heading is shown in the context in which it was used. The selected cards may easily be reproduced so that the requester can take a useful search product away from the file when the search is completed. **A** file system that uses this general approach of lifting sections from a single source document, reproducing it, and physically filing it under each of the categories or key words, has been referred to as a ***Collectanea.*** There are several examples of such systems in operation today.[18] Much credit for the initial development and subsequent improvement of this method must go to Igelsrud, Murdock, and Simpson at Battelle Memorial Institute. A good description of this method, as well as a discussion of its weak and strong points, has been given by Gibson and Lipetz.[19]

No. 31, JPRS: 8479 (June 22, 1961), distributed by the Office of Technical Services, U.S. Department of Commerce, Washington, D.C.

Gillie, P. J., "A Simplified Formula for Measuring Abstraction in Writing," *Journal of Applied Psychology,* Vol. 41, No. 4, pp. 214–217 (August 1957).

Rath, G. J., A. Resnick, and T. R. Savage, "Comparisons of Four Types of Lexical Indicators of Content," *American Documentation,* Vol. 12, No. 2, pp. 126–130 (April 1961).

[16] *Keywords Index to U.S. Government Technical Reports (Permuted Title Index),* Vol. 1, No. 4 (August 1,1962).

[17] This is essentially the same type of condensation done with some of the "automatic abstracting" computer programs.

[18] Hines, T. C., "The Collectanea as a Bibliographic Tool," Ph.D. Thesis, Rutgers University, 1961.

[19] Gibson, R. W., Jr., and B. Lipetz, "New Look in Manual Methods," *Special Libraries,* Vol. 47, No. 3, pp. 106–113 (March 1956).

Simpson, G. S., Jr., "The Qualitative Approach to Scientific Information," *Journal of Chemical Documentation,* Vol. 2, No. 2, p. 125 (April 1962).

Murdock, J. W., "Application of Battelle Technique to

KEYWORD INDEX

DISEASE	STUDIES ON THE IN VITRO MULTIPLICATION OF NEWC ASTLE DISEASE VIRUS IN CHICKEN BLOOD. III. A S TABILIZING SUBSTANCE FOR NEWCASTLE DISEASE VIR US PRESENT IN CHICKEN AND MAMMALIAN BLOOD CELL S PB 160 428(K) $1.60 0514
DISEASE	STUDIES ON THE IN VITRO MULTIPLICATION OF NEWC ASTLE DISEASE VIRUS IN CHICKEN BLOOD. III. A S TABILIZING SUBSTANCE FOR NEWCASTLE DISEASE VIR US PRESENT IN CHICKEN AND MAMMALIAN BLOOD CELL S PB 160 428(K) $1.60 0514
DISEASE	STUDIES ON THE IN VITRO MULTIPLICATION OF NEWC ASTLE DISEASE VIRUS IN CHICKEN BLOOD. I. VIRUS GROWTH IN RELATION TO AMOUNT AND KIND OF SEED VIRUS, TIME OF INCUBATION AND NUMBER OF CELLS PB 160 426(K) $2.60 0512
DISEASE	STUDIES ON THE MECHANISM OF IMMUNITY OF CHICKE NS TO NEWCASTLE DISEASE VIRUS. II. EXPERIMENTS CONCERNING THE MODE OF ACTION OF ANTIBODIES PB 160 431(K) $2.60 0507
DISEASE	STUDIES ON THE MECHANISM OF IMMUNITY OF CHICKE NS TO NEWCASTLE DISEASE VIRUS. I. INVESTIGATIO N OF THE POSSIBLE ROLE OF CELLULAR FACTORS PB 160 430(K) $2.60 0509
DISLOCATIONS	A STUDY OF THE PLASTIC DEFORMATION OF METALS B Y THE OBSERVATION OF SINGLE DISLOCATIONS AD-278 554(K) $6.60 0247
DISLOCATIONS	INTERACTION OF PRISMATIC AND GLISSILE DISLOCAT IONS UCRL-10213(K) $1.00 0174
DISLOCATIONS	THE EFFECT OF DISLOCATIONS ON DIFFUSION IN GER MANIUM NYO-10362(K) $1.60 0597
DISPLACEMENTS	ATOMIC DISPLACEMENTS IN METALLIC SOLID SOLUTIO NS TID-16488(K) $2.60 0214
DISPLAYS	LINEAL INCLINATION IN ENCODING INFORMATION SYM BOLICALLY ON CATHODE RAY TUBES AND SIMILAR DIS PLAYS AD-278 825(K) $1.75 0587
DISPOSAL	WASTE TREATMENT AND DISPOSAL PROGRESS REPORT F OR FEBRUARY AND MARCH 1962
DISSOCIATED	
	IONS FOR A LENTICULAR CYLINDER IN A SEMI-INFIN ITE FLUID AD-278 726(K) $4.60 0702
DISTRIBUTIONS	DIMENSIONS AND DISTRIBUTIONS OF CUMULUS CLOUDS AS SHOWN BY U-2 PHOTOGRAPHS AD-278 714(K) $6.60 0662
DISTRIBUTIONS	POPR-PROTOTYPE ORGANIC POWER REACTOR POWER DIS TRIBUTIONS NAA-SR-7413(K) $.50 0103
DISTRIBUTIONS	POWER DISTRIBUTIONS FOR TYPE 3 REPLACEMENT COR ES FOR SM-1, SM-IA AND PM-2A APAE-119(K) $10.50 0044
DISTRIBUTIONS	PRESSURE DISTRIBUTIONS ON TWO-DIMENSIONAL SHAR P-LEADING-EDGE FLAT PLATES WITH SWEEP ANGLES O F 0, 30, AND 45 AT A MACH NUMBER OF 6 AND ANGL ES OF ATTACK FROM 0 TO 90 NASA N62-15178(K) $.75 0456
DISTURBANCES	ON THE PROPAGATION OF AERODYNAMIC DISTURBANCES FAR FROM A BODY RISING THROUGH THE ATMOSPHERE AD-278 487(K) $4.60 0624
DIVERGENCE	DIVERGENCE CONDITIONS AND THE GRAVITATIONAL FI ELD EQUATIONS AD-278 433(K) $1.60 0685
DL	SYNTHESIS OF ORGANIC ACIDS BY SALMONELLA ANATU M FROM DL - ALANINE PB 160 451(K) $1.10 0525
DOCUMENTARY	MANUAL FOR OPERATIONAL DOCUMENTARY PHOTOGRAPHY TDO-10037(K) $.50 0098
DOCUMENTATION	THEORY OF DOCUMENTATION AND SEARCHING STRATEGY AD-278 551(K) $4.60 0742
DOG	TOLERANCE OF THE DOG TO EXTREME COLD EXPOSURE PB 160 378(K) $1.10 0533
DOGS	A TECHNIQUE FOR SAMPLING LYMPH IN UNANESTHETIZ ED DOGS BY MEANS OF AN EXTERIORIZED THORACIC D UCT-VENOUS SHUNT PB 160 368(K) $1.60 0491
DOGS	TRANSFUSION OF SEPARATE LEUKOCYTES ATED DOGS WITH ...

...I TO JULY 31,
...G. 1, 1962. CONTRACT
...K) $6.60

BIBLIOGRAPHY

... 30,
...K) $6.60

... ABSORPTION OF CARBON DIOXIDE AT
...URES, BY RICHARD H. TOURIN. REPT. ON
...RACT AF 18(600)596. MAY 54, 123P. 20 REFS.
TECHNICAL REPT. NO. 258, (OSR-TR-54-22), AD-37 571.
PB 160 934(K)$10.10

0739 WATERTOWN ARSENAL LABS., MASS.
FRACTURE TOUGHNESS OF AISI-4340 STEEL AS A FUNCTION OF PRIOR AUSTENITE GRAIN SIZE,' BY ROBERT N. KATZ. DEC 61, 12P. 5 REFS. TECHNICAL NOTE WAL TN 320.1/8.
PB 162 088(K) $1.60

0740 WATKINS-JOHNSON CO., PALO ALTO, CALIF.
DEVELOPMENT PROGRAM FOR MICROWAVE RADIATING POWER SOURCE FOR SPACE VEHICLE TELEMETRY TRANSMITTER. FINAL REPT. MAY 59-JULY 61, BY L. A. ROBERTS. JUNE 62, 64P. INCL. ILLUS. TABLES (CONTRACT AF 33(600)39392, PROJ. 4107) (ASD TR 61-711) AD-278 713(K) $2.00

0741 WESTERN ONTARIO U. (CANADA).
MOLECULAR EXCITATION GROUP. FINAL REPT. BY R. W. NICHOLLS. 1 JUNE 62, 34P. INCL. TABLES (CONTRACTS AF 19(604)1718 AND AF 19(604)4560) (AFCRL 62-699)
AD-278 511(K) $3.60

0742 WESTERN RESERVE U. CENTER FOR DOCUMENTATION AND COMMUNICATION RESEARCH, CLEVELAND, OHIO.
THEORY OF DOCUMENTATION AND SEARCHING STRATEGY. FINAL TECHNICAL DOCUMENTARY REPT. (1962) 41P. (CONTRACT AF 49(638)357) AD-278 551(K) $4.60

0743 WESTINGHOUSE ELECTRIC CORP., ELMIRA, N. Y.
AN INVESTIGATION OF GAS-FILLED THERMIONIC CONVERTERS FOR TRANSFORMING FISSION HEAT INTO ELECTRICITY, BY GENE R. FEASTER. INTERIM SCIENTIFIC REPT. NO. 1, 1 APR-30 SEP 60, ON CONTRACT AF 33(616)7198. 15 OCT 60, 177P. 113 REFS. AD-251. PB 162 021(K)$12.50

0744 WESTINGHOUSE ELECTRIC CORP., ELMIRA, N. Y.
FEASIBILITY OF THE USE OF HOLLOW CATHODES FOR THERMIONIC ENERGY CONVERSION. FINAL REPT. 1 MAR-30 NOV 61, BY I. BRODIE AND A. NIEWOLD. JUNE 62, 48P. INCL. ILLUS. TABLES, 29 REFS. (CONTRACT AF 33(616)7912, PROJ. 3145) (ASD TR 61-723) AD-278 824(K) $1.50

... AND EFFECTS OF MOISTURE TRANSLOCATION IN ...ED BAKED GOODS. PART I. PHYSICAL. PART II. BACTERIOLOGICAL, BY K. G. WECKEL AND ELIZABETH MCCOY. REPT. NO. 9 (FINAL), 15 JAN 59-14 JULY 61, ON CONTRACT DA 19-129-QM-1362. (1961) 50P. 25 REFS. FILE NO. C-322.
PB 160 944(K) $4.60

0755 WISCONSIN U. MATHEMATICS RESEARCH CENTER, MADISON.
A BOUNDARY LAYER THEORY FOR UNSYMMETRIC DEFORMATIONS OF CIRCULAR CYLINDRICAL ELASTIC SHELLS, BY M. W. JOHNSON, JR. MAY 62, 54P. 24 REFS. (MRC TECHNICAL SUMMARY REPT. NO. 318) (CONTRACT DA 11-022-ORD-2059)
AD-278 744(K) $5.60

0756 WISCONSIN U. MATHEMATICS RESEARCH CENTER, MADISON.
A SET OF MATRICES FOR TESTING COMPUTER PROGRAMS, BY J. L. BRENNER. MAY 662, 7P. (MRC TECHNICAL SUMMARY REPT. NO. 325) (CONTRACT DA 11-022-ORD-2059)
AD-278 733(K) $1.10

0757 WISCONSIN U. MATHEMATICS RESEARCH CENTER, MADISON.
MULTIPLE SCATTERING OF PLANE WAVES BY TWO ELLIPTIC CYLINDERS, BY N. ZITRON. MAY 62, 31P. INCL. ILLUS. 21 REFS. (MRC TECHNICAL SUMMARY REPT. NO. 316) (CONTRACT DA 11-022-ORD-2059) AD-278 745(K) $3.60

0758 WISCONSIN U. MATHEMATICS RESEARCH CENTER, MADISON.
NOTE ON DIFFERENCE APPROXIMATIONS WITH NON-NEGATIVE COEFFICIENTS FOR PARTIAL DIFFERENTIAL EQUATIONS, BY DONALD GREENSPAN. MAY 62, 16P. 3 REFS. (MRC TECHNICAL SUMMARY REPT. NO. 323) (CONTRACT DA 11-022-ORD-2059)
AD-278 743(K) $1.60

0759 WISCONSIN U. MATHEMATICS RESEARCH CENTER, MADISON.
SOME EXTREMAL QUESTIONS FOR SIMPLICIAL COMPLEXES. III. PROBLEMS OF THE GEOMETRY AND ANALYSIS OF THE HIGHER EUCLIDEAN SPACES, BY L. C. YOUNG. MAY P. 15 REFS. (MRC TECHNICAL SUMMARY REPT. NO. 320) (CONTRACT DA 11-022-ORD-2059) AD-278 741(K) $3.60

0760 WISCONSIN U. MATHEMATICS RESEARCH CENTER, MADISON.
SOME EXTREMAL QUESTIONS FOR SIMPLICIAL COMPLEXES. IV. THE ALGEBRAIC AND THE GEOMETRIC RESULTANT, AN APPLICATION OF VARIATIONAL METHODS, BY W. H. FLEMING AND L. C. YOUNG. MAY 62, 19P. 14 REFS. (MRC TECHNICAL SUMMARY REPT. NO. 322) (CONTRACT DA 11-022-ORD-2059)
AD-278 742(K) $2.60

0761 WISCONSIN U. MATHEMATICS RESEARCH CENTER, MADISON.
THE ELASTODYNAMICS OF MOVING LOADS. PART II. THE STRESS AND VELOCITY FIELD OF AN ENERGY SOURCE MOVING IN AN ELASTIC SOLID WITH A FREE SURFACE AT CONSTANT VELOCITY, BY MICHAEL PAPADOPOULOS. MAY 62, 30P. 5

FIG. 2-3 Representative sample of a permutation or Key-Word-in-Context index.

Ti 48209 Card 1/2

J. D. Jackson, P. D. Miller, W. K. Boyd, F. W. Fink
A Study of the Titanium-Liquid Oxygen Pyrophoric Reaction
Battelle Memorial Institute, Columbus, Ohio, WADD TR 60-258, June, 1960,
(June 1, 1959 - February 15, 1960), AF 33(616)-6345)

...
Early indications of the reactivity of titanium with LOX stimulated studies by several aircraft and other interested companies. Most of these early investigations were on impact sensitivity. Impact sensitivity is determined by a reaction with LOX upon impact by a falling weight. Titanium has exhibited impact sensitivity at a lower energy level than is accepted for other metals, that is, no reactions out of 20 impacts at the 70-foot-pound level. Titanium alloys were found more reactive than unalloyed titanium. Alloys in the aged or hardened form were found most reactive.

Other investigators indicated that, when titanium was fractured under LOX to present a fresh surface, a reaction occurred.

The physical properties of titanium would suggest a higher reactivity to LOX than in the case with other metals. For example, titanium has (1) low thermal conductivity and law heat capacity on a volume basis, resulting in high temperature and slow heat dissipation for a given heat input; (2) a low ignition temperature in an oxygen atmosphere; and (3) a reactivity in high-pressure oxygen at and above room temperature.

Ti 48209 Card 2/2

The experimental program at Battelle was designed to investigate important factors singly and in a controllable manner in the titanium-LOX reaction. The following factors have been studied in detail using titanium that was thoroughly cleaned:

(1) Exposure under LOX of a fresh titanium surface by fracture or tearing
(2) Localized impact of a surface under LOX to cause deformation
(3) Galling of the surface under LOX with a scribe under a heavy load
(4) Impact of clean smooth surfaces under LOX.

It was concluded from the above experiments that these factors were not a primary cause of the reaction. It appears that a combination of several factors in proper sequence is necessary to initiate the reaction. A proposed mechanism is that gaseous oxygen is formed by the heat generated at local impact sites, this trapped oxygen is compressed, and the freshly exposed titanium metal reacts locally with the high-pressure oxygen. Propagation is contingent on the heat generation and the quenching capacity of the system.

...

[53 pages, 17 figures, 8 tables, 23 references - Disposition, Shelf]
JM-msf September 25, 1962

FIG. 2-4 Representative clue word or extract cards.

Words Chosen from Text, Omission of Common Words, and Consideration of Variants

An improvement can be realized over the indexing method just described by incorporating some additional terminology control to take care of all the variant word forms used for indexing; for example, a list of synonyms (with a "use" or "see" reference) to indicate which terms to use. Another form of control is the use of definitions or brief descriptions for the words that are spelled the same, but have different meanings, such as: (1) pitch (to throw); (2) pitch (tar); (3) pitch (acoustics); (4) pitch (attitude or angle). Figure 2-5 illustrates this characteristic. Words which have common roots or are variants of the same basic term can be related (e.g., fungus, fungi; sun, solar; printing, printed).

Words Chosen from Text, Omission of Common Words, and Consideration of Variants and Generic Relationships

A further improvement is obtained by the determination and display of generic relationships between the indexing terms (a "generic to" and "specific to" reference). Once these relationships are established for terms already in the indexing system, there is a tendency to standardize the indexer's choice of terms chosen from new documents. As this list is examined, relationships between terms are noticed and recorded. An example of such a display is the ASTIA thesaurus (a portion is shown in Fig. 2-5). A thesaurus is essentially an alphabetical list of possible indexing terms, each of which is accompanied, where relevant, by its related terms, references, and scope notes. It is an authority list which shows relationships of terms, and helps to bring terms to the attention of the indexer and the searcher. Several comprehensive technical thesauruses or authority lists of this type have been published recently—by ASTIA, the American Institute of Chemical Engineers, and the U.S. Atomic Energy Commission.[20] These lists assist both the indexers and searchers of scientific and technical material.

To date, most of the reported dissatisfaction with Uniterm indexing has fallen into seven different categories. As described by Costello, these categories are: (1) semantic problems; (2) viewpoint; (3) generics, or specificity and generality of description; (4) Syntactical controls and potential false retrievals, or noise; (5) generalities as indexing terms; (6) ranges of numerical information; and (7) posting.[21] The thesaurus or authority list approach gives indications of being able to solve some of the problems of generics, viewpoint, and semantic~.[22] Preparation of a good thesaurus requires a great deal of effort, perhaps several man-years, but it is a necessity for any efficient file system.

Words Chosen from Text and Consideration of Syntactical Relationships between Indexing Terms

In addition to all the vocabulary control measures taken in the preceding methods, the indexing system could also show the functional or syntactical relationships between the indexing terms used to describe the content of a single document. For example, the indexing terms man, bites, and dog appearing in an index would not distinguish whether a document indexed under these three terms was about a dog biting a man, or vice versa. Some of this ambiguity can be reduced by assigning role indicators to the index terms to show the role or application of each term, as well as to indicate which terms are

the Operation of Information Centers," *Journul of Chemical Documentation,* Vol. 2, No. 2, pp. 126127 (April 1962).

Olofson, C. T., "Mechanics of Answering a Technical Inquiry," *Journal oj Chemical Documentation,* Vol. 2, No. 2, pp. 127–128 (April 1962).

[20] Heald, J. H., et al., *Thesaurus of ASTIA Descriptors,* 2nd ed. (Armed Services Technical Information Agency, Arlington Hall, Virginia, 1962).

***A.I.Ch.E. Chemical Engineering Thesaurus* (American Institute of Chemical Engineers, New York, 1961).**

***Subject Headings Used by the USAEC Technical Information Service,* C. W. Hargrave, editor, Report TID-5001 (4th rev.) of the United States Atomic Energy Commission, Division of Technical Information (August 1962). The original edition was published in July 1951.**

[21] Costello, J. C., "Uniterm Indexing Principles, Problems, and Solutions," *American Documentation,* Vol. 12, No. 1, pp. 20–26 (January 1961).

[22] Joyce, T., and R. M. Needham, "The Thesaurus Approach to Information Retrieval," *American Dmumentatwn,* Vol. 9, No. 3, pp. 192–197 (July 1958).

Holm, B. E., end L. E. Rasmussen, "Development of a Technical Thesaurus," *American Documentation,* Vol. 12, No. 3, pp. 184–190 (July 1961).

Vickery, B. C., "Thesaurus—A New Word in Documentation," *Journal of Documentation,* Vol. 10, No. 4, pp. 181–189 (December 1960).

Frick, B. M., "Suggestions for the Beginner in Subject Heading Work," an introductory chapter in *Sears List of Subject Headings,* 8th ed., by B. M. Frick (H. W. Wilson Co., New York, 1959).

EAR - EDV

E

EAR
  (ANATOMY)
  INCL: COCHLEA
    LABYRINTH
    SEMICIRCULAR CANALS
    STAPES
    VESTIBULAR APPARATUS
  ALSO SEE: OTORHINOLARYNGOLOGY

EAR MUFFS USE EAR PROTECTORS

EAR PROTECTORS
  (SAFETY DEVICES AND PROTECTIVE EQUIPMENT)
  INCL: EAR MUFFS

EAR WAX USE CERUMEN

EARLY WARNING RADAR
  (RADAR AND RADIO DETECTION)
  ALSO SEE: SEARCH RADAR

EARLY WARNING RADAR SYSTEMS USE EARLY WARNING
  RADAR AND RADAR EQUIPMENT

EARLY WARNING SONAR
  (ACOUSTIC DETECTION)

EARPHONES
  (COMMUNICATION SYSTEMS AND COMPONENTS)
  ALSO SEE: HEARING AIDS
    HEADPHONES

EARTH
  (ASTRONOMY)
  ALSO SEE: ATMOSPHERE
    GEOCHEMISTRY
    GLACIERS
    GRAVITY
    LITHOSPHERE
    PLANETS
    ROCK
    SOILS
    SOLAR SYSTEMS
    WATER
    SPECIFIC TYPES OF LAND FORMS OR BODIES
      OF WATER, E.G.: MOUNTAINS; OCEANS

EARTH AUGERS
  (CONSTRUCTION EQUIPMENT AND MATERIALS)
  ALSO SEE: DRILLS

EARTH MODELS
  (GEOPHYSICAL CONDITIONS)

EARTH MOVING EQUIPMENT
  (CONSTRUCTION EQUIPMENT AND MATERIALS)
  INCL: SOIL REMOVAL
  ALSO SEE: BULLDOZERS
    GRADERS
    ROAD BUILDING EQUIPMENT

EARTH SLIDES USE AVALANCHES

EARTHQUAKE-RESISTANT STRUCTURES
  (STRUCTURES (BUILDINGS AND CIVIL ENGI-
  NEERING))

EARTHQUAKE WARNING SYSTEMS
  (SEISMOLOGY)

EARTHQUAKES
  (SEISMOLOGY)
  ALSO SEE: DISASTERS
    SEISMIC WAVES
    SEISMOLOGICAL STATIONS

EAST CHINA SEA
  (GEOGRAPHY)

EAST GERMANY
  (GEOGRAPHY)
  ALSO SEE: EUROPE

EASTERN HEMISPHERE
  (GEOGRAPHY)

EBERTHELLA USE SALMONELLA
EBUR USE DENTINE

ECHINODERMS
  (MARINE BIOLOGY)
  INCL: HOLOTHURIANS
    SEA CUCUMBERS
  ALSO SEE: AQUATIC ANIMALS

ECHO RADAR MARKERS USE DELAYED ECHO RADAR MARKERS

ECHO RANGING
  (WAVE PROPAGATION)
  ALSO SEE: SOUND RANGING

ECHO RANGING TORPEDOES
  (UNDERWATER ORDNANCE)
  ALSO SEE: ACOUSTIC TORPEDOES
    HOMING TORPEDOES

ECITON USE ANTS
ECLIPSES USE LUNAR ECLIPSES; STELLAR ECLIPSES

ECOLOGY
  (BIOLOGY)
  INCL: BIONOMICS
  ALSO SEE: BIOLOGY
    CLOSED-CYCLE ECOLOGICAL SYSTEMS
    PALEOECOLOGY

ECONOMIC ASPECTS USE ECONOMICS

ECONOMIC CONDITIONS
  (ECONOMICS)
  ALSO SEE: ECONOMICS

ECONOMIC MOBILIZATION
  (LOGISTICS)

ECONOMIC WARFARE
  (LOGISTICS)

ECONOMICS
  (ECONOMICS)
  INCL: ECONOMIC ASPECTS
  ALSO SEE: ECONOMIC CONDITIONS
    SOCIAL SCIENCES

ECUADOR
  (GEOGRAPHY)

EDDY CURRENT WAKES USE MAGNETIC BRAKES
EDDY CURRENT CLUTCHES USE MAGNETIC CLUTCHES

EDEMA
  (PATHOLOGY)
  INCL: OEDEMA

EDIE'S TOLL PROBLEM USE SCHEDULING
EDISON CELLS USE ALKALINE CELLS

EDUCATION
  (SOCIAL SCIENCES)
  INCL: ILLITERACY
    LITERACY
    TEACHING
  ALSO SEE: CULTURE
    INSTRUCTORS
    LEARNING
    READING
    STUDENTS
    TEXTBOOKS
    TRAINING
    UNIVERSITIES

EDUCATIONAL DYNAMICS USE GROUP DYNAMICS
EDUCTORS USE AIR EJECTORS; EXHAUST GAS EJECTORS;
  FUEL EJECTORS; JET PUMPS; PUMPS
EDVAC USE DIGITAL COMPUTERS

135

FIG. 2-5 Portion of representative indexing thesaurus.

being acted upon by the other terms, and in what manner.

With regard to Uniterm systems, these special tags have been referred to in the literature as "links" and "roles."[23] "Links" indicate the fact that terms *are* related, but do not indicate how. A "role" is appended to each term in a link to indicate the function of that term. Thus, in multiple-term associations, roles indicate *how* terms are related to each other. The indexing task becomes more complex and difficult when links and roles are used or when "modifier lines" need to be composed. Great care must be exercised by the indexers.

Another method of syntactical control is to maintain the integrity of the phrase which identifies the subject content of the document. A good example of the use of this method is the subject index to *Nuclear Science Abstracts.*

The telegraphic abstracting system of Western Reserve University is similar in many respects to the links and roles scheme in the sense that it uses role indicators and other means for showing the functional relationship between the various indexing components.

Any of Preceding Methods, with Addition of Terms Not Used in the Text

The quality of the indexing of any of the previously mentioned schemes can be improved by adding index terms that are implied but not actually stated in the text, or arranging phrases in an inverted order for more accessible index points. This step, requiring judgment and some subject knowledge, marks the first clear-cut point where derived indexing starts to change to assigned indexing. Up to this point, all of the indexing methods could be performed almost automatically by people or machines—using such tools as tables of allowable index words and rules for systematically extracting words from the text. Up to this point, the index terms could be derived from the document to be indexed by following specified rules or algorithms while examining the text. The automatic indexing systems have in fact achieved success in implementing almost all parts of the first four methods described in this section. There will, however, be difficulty in achieving high-quality machine indexing at or above this fifth level of indexing complexity. Nevertheless, it seems possible that machine methods may be used to generate most of the indexing data for these more complex indexing schemes, with some additional indexing data assigned by a human indexer. Some recent automatic indexing methods do in fact take this approach, with the final product being the result of combined machine and manual effort.[24] The indexer generates new terms, or arrangements of terms, or the thesaurus or authority lists mentioned in previous sections are used for the additional index terms not found in the document.

Assignment of Index Entries from a Fixed Authority List or Classification Scheme

The most carefully controlled index language is usually found in those systems which use a fixed list of index terms or classification scheme. New terms or class numbers are added reluctantly. In such systems the indexer must use the subject headings or terms that appear in the authority list. Revision of the authority for the terms or class numbers is possible, but often the procedure or mechanism for the revision is unwieldly and hampered by long delays. Examples of such fixed index languages are the familiar Dewey Decimal classification and the Library of Congress List of Subject Headings.[25] Each classification scheme that uses a

[23] Taube, M., "Notes on the Use of Roles and Links in Coordinate Indexing," *American Documentation,* Vol. 12, No. 2, pp. 98–100 (April 1961).

Costello, J. C., Jr., "Storage and Retrieval of Chemical Research and Patent Information by Links and Roles in DuPont," *American Documentation,* Vol. 12, No. 2, pp. 111–120 (April 1961).

Costello, J. C., Jr., "Some Solutions to Operational Problems in Concept Coordination," *American Documentation,* Vol. 12, No. 3, pp. 191–197 (July 1961).

Costello, J. C., Jr., "Uniterm Indexing Principles, Problems, and Solutions," *American Documentation,* Vol. 12, No. 1, pp. 20–26 (January 1961).

Montague, B. A., "Patent Indexing by Concept Coordination Using Links and Roles," *American Documentation,* Vol. 13, No. 1, pp. 104-111 (January 1962).

[24] Borko, H., "The Construction of An Empirically Based Mathematically Derived Classification System," *Proceedings of the 1962 Spring Joint Computer Conference,* San Francisco, May 1962, pp. 279–289.

Swanson, D., *Word Correlation and Automatic Indexing; Phase IIA Final Report—Appendix A: Thesaurus; Appendix B: Alphabetic Index to Thesaurus* (Thompson Ramo Wooldridge, Inc., RW Division, Canoga Park, California, 1962).

[25] Dewey, M., *Dewey Decimal Classification and Relative Index,* 16th ed. (Forest Press, Inc., Lake Placid Club,

numeric or alphanumeric notation for classified arrangement has a well-developed and standardized descriptor language to accompany it.[26] In some cases these languages have been developed and maintained over a period of 60 years or more.

Such well-developed languages may or may not have a generic or hierarchical structure built into them. The Library of Congress subject heading list has been shown, for example, to contain a hidden classification structure.[27]

One other system with a well-controlled language, but without a hierarchical structure is Mooers' Descriptor system. In the early 1950's, Mooers used the word "descriptor" to describe the specific type of indexing or subject terms that he was using.[28] This "descriptor" (which we shall refer to as a Descriptor with a capital D) has a special and well-defined meaning, although common usage in the documentation field considers "descriptor" to be synonymous with "Uniterm," "key word," and "subject heading." The Descriptor is a broad subject heading that stands for an idea or concept, and is chosen to facilitate retrieval for a particular group of users. That is, the Descriptors are tailored to a specific population of users. Many other indexing systems that have little control of the index language may develop an authority list of perhaps 10,000 terms. However, the Descriptor systems use a very tightly controlled indexing vocabulary of perhaps 250 to 350 Descriptors. The schedule of Descriptors is separately derived and maintained on an empirical basis for each file collection with its particular population of users—thus, in theory, providing access to the material from the same point of view as that of the file user. A dictionary or thesaurus is established for each system, with many "see" references and definitions for each Descriptor. "See" references may be made from words in the users' own terminology (as determined by the indexer) to locate the appropriate Descriptors. Some observers have suggested that the Descriptor system is essentially the same as the Uniterm system, with the Descriptors being used to code the Uniterms.[29]

Assignment of Index Entries from Authority Lists or Classification Schemes Representative of Several Viewpoints and *Aspects* of Subject

An indexing or classification scheme usually represents one particular viewpoint of the persons or persons who developed it. A classification schedule or a list of descriptors may organize a body of information into some very useful form for one user, but may not provide the type of information needed by another user with another viewpoint. For example, a cocker spaniel might logically be found in a classification scheme for the animal kingdom; but he might also be considered from a different viewpoint when placed in a classification scheme for various types of burglar alarms. This same dog could be represented as one part of an infinity of different organizations of a collection of information.

There is no single way to organize the entire universe of information into an ordered and well-structured pattern of relationships that is best for all users. Most methods of file organization represent a single approach or point of view. For maximum utility, any single approach should follow the methods of organization and division of material used consciously or unconsciously by most of the users of that particular file.

Fortunately, it is not necessary to restrict the indexing of a particular item to a single approach or point of view. It is quite possible to analyze the elements and characteristics of a file item from several points of view and synthesize a group of indexing terms for the item. This is often called "facet analysis" or "relational indexing." It

New York, 1958). See also Guide to the Use of Dewey Decimal Classification (Forest Press, 1962). The first edition was published about 1876.

Library of Congress, Classification, Class T (Technology), 4th ed. (1953). The first edition was published in 1903.

[26] Frick, B. M., Sears List of Subject Headings, 8th ed. (H. W. Wilson Co., New York, 1959).

Subject Headings Used in the Dictionary Catalogs of the Library of Congress, 6th ed., M. V. Quattlebaum, editor (Library of Congress, Washington, D.C., 1957). This 6th edition contains the headings established and applied by the Library of Congress from 1897 through 1955. The 1st edition was published in sections from 1910 to 1914. A 7th edition is now in preparation.

[27] Richmond, P. A., "Cats: An Example of Concealed Classification in Subject Headings," Library Resources and Technical Services, Vol. 3, No. 2, pp. 102-112 (Spring 1959).

[28] Mooers, C. N., "Zatocoding and Developments in Information Retrieval," Association of Special Libraries and Information Bureaux Proceedings, Vol. 8, No. 1, pp. 3–22 (February 1956).

[29] Jaster, J. J., "A Note on Descriptors," American Documentation, Vol. 13, No. 4, pp. 433-434 (October 1962).

amounts to the composite indexing of an item by combining in a prescribed manner the terms derived from separate indexing examinations. A "synthetic" classification schedule, and index terms, or point of view symbols (i.e., multiple facets) are derived for each aspect of the subject.

Ranganathan's Colon classification[30] and the recent work of members of the Classification Research Group in England[31] are outstanding examples of this method of indexing and classifying. Ranganathan's first work on the subject was published in 1933 and is still being expanded. These analytic and synthetic methods provide a systematic specification of complex subjects. According to Bliss, Ranganathan's work is a validation of the principle of composite classification, systematic and synthetic.[32] Instead of providing a completed classification schedule, or list of terms, Ranganathan has outlined a method and a systematic plan whereby each user can construct his own classification schedule and index language. In this way, a file item receives its composite indexing by the assembling, in a prescribed sequence, of all the terms and code groups. Lists are empirically derived for each installation, and can be expanded. The final code designation for the file item consists of a sequence of code groups separated by colons, with the various aspects arranged in sequence according to their importance to the user. The code groups can show generic relationships. Selection by a single aspect is also possible. By first grouping the terms into categories or facets, and providing a method of notation to link the facets together, the classification scheme displays the index terms together in a coordinate relationship to designate a complex idea. By providing for a "chain index," each aspect of the subject becomes an index term.

Although the work of Ranganathan[33] has had considerable impact on the recent work in classification, there are few installations actually using a complete Colon classification system. The Metal Box Co., Ltd., in England uses the Colon classification, in part, to classify their books and reports on packaging.

Some of the facet analysis techniques have recently been adapted by the British National Bibliography. For a more complete description, and a historical survey of the development of Colon classification, the reader is referred to articles by Tyaganatarajan and Foskett.[34]

The Semantic code, another type of faceted classification, developed by Perry, Kent, and others at Western Reserve University around 1950, is a method for coding information for machine searching and is composed of a number of related index terms known as "semantic factors," each of which represents a general conceptual area of the word being encoded.[35]

[30] *Ranganathan, S. R., Colon Classification, 6th ed. (G. Blunt & Sons, London,* 1951). *[So called because a colon (:) is used in the notation to separate the index terms of each* facet.]

Ranganathan, S. R., "Telescoping of Facets and Mixed Notation," Revue de la Documentation, Vol. 27, *No.* 4, *pp.* 154–156 *(November* 1960).

Ranganathan, S. R., *"Classifying, Indexing, Coding," Chapt.* 3 in *Information Retrieval and Machine Translation, Part I, A. Kent, editor* (Interscience *Publishers, New York,* 1960).

Ranganathan, S. R., Elements of Library Classification, 2nd ed. (The Association of Assistant Librarians, London, 1959).

Ranganathan, S. *R., Prolegomena to Library Classification, 2nd ed. (Library Association, London,* 1957).

Ranganathan, S. R., "Colon Classification and Its Approach to Documentation," in Bibliographic Organization, J. Shera, editor (University of Chicago Press, 1951).

[31] *Foskett, D. J., et al., "Classification Research Group Bulletin No.* 7," *Journal of Documentation, Vol.* 18, *No.* 2, *pp.* 65-88 *(June* 1962).

Vickery, B. C., "Developments in *Subject Indexing," Journal of Documentation,* Vol. 11, *No.* 1, *pp.* 1–11 *(March* 1955).

Vickery, B. C., Faceted Classification: A Guide to Construction and Use of Special Schemes (Association of Special Libraries and Information Bureaux, London, 1960).

[32] *Bliss,* H. *E., Organization of Knowledge in Libraries, 2nd ed., p. 304* (H. *W. Wilson Co., New York,* 1939).

[33] *Ranganathan, Colon Classification, cited earlier, fn.* 30.

[34] *Tyaganatarajan, T., "A Study in the Developments of Colon Classification," American Documentation, Vol.* 12, *No. 4, pp.* 270–278 *(October* 1961).

Foskett, D. J., "The Construction of a Faceted Classification *for a Special Subject," Proceedings of the International Conference on Scientific Information, Vol.* 2, *pp.* 867–888 *(National Academy of Sciences—National Research Council, Washington, D.C.,* 1959).

[35] Melton, J. *L., "The Semantic Code," pp.* 221–279 *in Tools for Machine Literature Searching, J. W. Perry and A. Kent, editors (Interscience Publishers, New York,* 1958).

Melton, *J. L., "The Semantic Code Today," American Documentation, Vol. 13, No.* 2, *pp.* 178–181 *(April* 1962).

Taube, *M., "A Note on the Evaluation of the WRU Semantic Code as an Example of Generic Coding," American Documentation, Vol.* 13, *No.* 2, *pp.* 185–186 *(April* 1962).

Vickery, B. C., "The Structure of 'Semantic Coding': A Review," American Documentation, Vol. 10, *No.* 3, *pp.* 234–241 *(July* 1959).

The Semantic code is similar to the Colon classication, with the differences being predominantly in the notation.[36] This is due in large part to the fact that the Colon classification was designed for visual consultation, and the Semantic code (called Encoded Telegraphic Abstracts) was designed for machine searching. There seems to be a high degree of compatibility between the two systems. The Semantic code does not rely entirely on words contained in the text of the file item. During its early development, the code was based on terminology drawn primarily from classification systems and subject-heading lists from all fields of science.[37] The coding system began to be used in 1956 to encode metallurgical abstracts for the American Society for Metals and the terminology of the code became rather heavily oriented toward metallurgy and immediately adjacent subject fields.[38]

In early 1961 the Semantic code dictionary contained over 35,000 terms (over 19,000 terms encoded with the Semantic code, and over 16,000 chemicals, names, etc., encoded with non-Semantic codes). To date, the Semantic codes have been applied primarily to the area of physical sciences, although some studies have been made of its application to such fields as law and medicine. The Semantic coding system forms the basis for an operational literature searching service currently being operated at Western Reserve University under the sponsorship of the American Society for Metals.[39] This coding system is not used widely—its application so far has been primarily limited to the research group at Western Reserve University.

[36] Melton, J., *A Note on the Compatibility of Two Information Systems, Colon Classification and Western Reserve University (Encoded Telegraphic Abstracts) and the Feasibility of Interchanging Their Notations, Technical Note No. 13 of the Documentation and Communication Research Center, Western Reserve University, Cleveland, Ohio (1960); Air Force Officeof Scientific Research Technical Note TN 60-261 on Contract AF 49(638)-357.*

[37] Melton, J. L., *Test Program for Evaluating Procedures for the Exploitation of Literature of Interest to Metallurgists. V. The Semantic Code Today, a report of the Center for Documentation and Communications Research at Western Reserve University, Cleveland, Ohio (1961).*

[38] Perry, J. W., A. Kent, and M. M. Berry, *Machine Literature Searching (Interscience Publishers, New York, 1956).*

[39] *Kent, A.,* J. Melton, *and C. Flagg, "Abstracting, Coding, and Searching the Metallurgical Literature for A.S.M. The WRU Searching Selector," Chapt. 23 in Information Systems in Documentation, J. H. Shera et al., editors (Interscience Publishers, New York, 1957).*

The Universal Decimal Classification (UDC) also has a facet device (:) and auxiliary point-of-view indicators to add to the main UDC number to suggest a particular point of view of the document.[40] Separate schedules or classifications are available for such points of view as: theoretical (aim, trials and tests, research, and development); practical (realization, execution, production, materials, plants, products); economic and commercial; use, operation, and care; installation; personnel; organization or management; and social or ethical (responsibilities, obligations).

METHODS OF SUBJECT INDEX DISPLAY

There are many ways to display the file index to its users (from catalog cards to printed indexes) with each method of display providing different operational features, even though the same method or degree of indexing is used. The choice of display technique can be almost independent of the choice of the indexing system, and the same indexing system can be displayed in many different forms. Thus, the designer of the file system has the freedom to choose the indexing system, notation, and display technique almost independently. When regarded in this light, we see that many "different" indexing systems are actually variations in the form of display, with no basic change in the power or capability of the indexing system itself. A number of different methods of display are described briefly below. Equipment and methods for preparing these different forms of display are described in later chapters of this book.

Unit-Record Displays

A very common approach is to prepare a single card (a unit-record) for each file item, and to duplicate and file a copy of that card under each heading or subject term that is used to describe that document. This approach is used with conventional library card catalogs, with the cards (see Figs. 2-6 and 2-7) usually filed by the classificgtion number, author(s), and subject headings. The record of where the duplicate cards have been filed is indicated in the "tracing," the numbered terms at the

[40] *Universal Decimal Classification, Abridged English Edition, 3rd ed., revised 1961 (British Standards Institution, London, 1961). The first English edition was published by the International Federation for Documentation in Brussels in 1927-1933.*

The National union catalog; a cumulntive author list representing Library of Congress printed cards and titles reported by other American libraries. 1956–
Washington, Library of Congress.
v. 28 cm.
Printed in 9 monthly issues, 3 quarterly cumulations, annual cumulations for four years and a quinquennial in the fifth.
In 1958 a cumulation, covering 1956–57 entries and entries in the 1958–55 annual vols. of the Library of Congress catalog ... —Books: authors, was published in Ann Arbor, Mich. by J. W. Edwards.
At head of title, 1956– The Library of Congress catalogs.
Supersedes Library of Congress catalog; a cumulative list of works represented by Library of Congress printed cards—Books: authors.
Compiled by the Processing Dept. ot the Library of Congress (with the cooperation of the Committee on Resources of American Libraries
(Continued on next card)
56—60041
[60r59n⁴15]

The National union catalog ... (Card 2)
(called July 1956–Oct. 1957 Board of Resources of American Libraries) of the American Library Association, July 1956–
Includes entries for map and atlases.
Quinquennial cumulations include the quinquennial cumulations ot Library of Congress cntnlog: Music and phonorecords, and Library of Congress catalog: Motion pictures and filmstrips.

1. Catalogs, Union. 2. American literature—Bibl.—Catalogs. I. U. S. Library of Congress. Processing Dept. II. Title: The Library of Congress catalogs.
Z881.A1U372 018.1 56—60041
——— 3d set. Z663.7.L512
Library of Congress [60r59n⁴15]

FIG. 2-6 Representative Library of Congress catalog card.

CID 1352 USSR
Ryskin, G. I͡A.
DIFFUSION COEFFICIENT MEASUREMENT BY RADIOACTIVATION ANALYSIS OR THE ISOTOPIC DILUTION METHOD. (Izmerenie koeffit͡sientov diffuzii metodami radioaktivat͡sionnogo analiza i izotopnogo razbavleni͡a; Text in Russian.) Fizika Tverdogo Tela, 1, no. 6: 952-954, June 1959. 6 refs.
DLC, Unbound periodical
In case of very small diffusion coefficients ($D \cong 3 \times 10^{-17}$ cm/sec), the customary methods of diffusion analysis contain many errors. In these cases, activation or isotope dilution analyses give more exact results. The first method consists in the preparation of the diffusing metal and the diffusion medium, the annealing of both, the separation of the metal and medium, the bombardment of the medium
(over)

1. Metals--Diffusion --Measurement I. Ryskin, G. I͡A. II. Akademi͡a Nauk SSSR. Leningradskiĭ Fiziko-Tekhnicheskiĭ Inst. [Acad. Sci. USSR. Leningrad Inst. Physics and Technology]
U.S. Atomic Energy Commission
Office of Isotopes Development

FIG. 2-7 Representative side-margin catalog card.

bottom of Fig. 2-6 and on the right side of Fig. 2-7. Extensive use of the conventional library catalog card is illustrated by the fact that in 1962 the Library of Congress sold over 42 million printed cards for use in other libraries, and filed over 1½ million cards in its own library. Reference cards ("see" and "see also") are considered part of the display. This unit-record approach is also used with the clue-word or extract systems, with the cards filed under terms that appeared in the text (see Fig. 2-4). One modification of this unit-record approach is to print, in book form, copies of the cards in the order that they would appear in the card drawer file. One of the Library of Congress publications is prepared in this manner, by photographing the cards and printing them in reduced size, 18 to a page.[41]

Unit-Term Displays

Another approach, used primarily with coordinate indexing systems, is to prepare a single record (usually a single sheet of paper) for each subject heading or index term, and record on each term card the names or numbers of every file item indexed by that particular term (see Fig. 2-1). Most "draft" copies of published indexes take this form, regardless of what method of indexing has been used. One modification of this technique is to bind and publish copies of all the term cards together under one cover. This creates another type of book-form index, several more of which are described below.

Book-Form or "Published Index Displays

Arranged by Text Words. Several types of indexes are prepared in book form and organized by the words used in the text material, rather than by a pre-determined classification scheme. Figure 2-8 illustrates two portions of a printed book catalog that was prepared to provide a condensed description of the holdings of an engineering library, and meant to be used as a supplement to the library's regular card catalog.[42] Notice that the one-line entries are sorted and listed by author and by title, but not by the classification number. Thus the index is organized and displayed on the basis of the descriptive text material. The index could be prepared either manually or with the aid of some data processing equipment. In this case, punched card equipment was used. Notice that this form of display, completely different from the regular card catalog, and with an intended use different from the card catalog, has the same classification information as the card catalog, namely, the Library of Congress

[41] *A Catalog of Books Represented by Library of Congress Printed Cards Issued to July 31, 1942 (Edward Bros., Ann Arbor, Michigan,* 1942–1946).

[42] *Richmond,* P. *A., "A Short-Title Catalog Made with IBM Tabulating Equipment," Library Resources and Technical Services, Vol.* 7, *pp.* 81–90 *(Winter* 1963).

MASTER CATALOG ARRANGED BY TITLE

FIELD ENGINEERING ED. 22	57	SEARLES W H	TF0205
FIELD THEORY OF GUIDED WAVFS	60	COLLIN ROBERT E	QC0661
FIFTY YEARS OF PAINT TESTING	53	AMER SOC TEST MAT	TP0936.5
FILTRATION & FILTERS	29	PICKARD JOSEPH ALLEN	TP0156
FINANCIAL STATEMENT ANALYSIS	41	MYER JOHN N	HF5681
FINE PARTICLE MEASUREMENT	59	ORR CLYDE	TA0406.7
FINISHING METAL PRODUCTS ED. 2	46	SIMONDS HERBERT R	TS0213
FINITE DEFORMATION OF AN ELASTIC SOLID	51	MURNAGHAN FRANCIS D	QA0935
FINITE DIFFERENCE METH PART DIFF EQUATIONS	61	FORSYTHE GEORGE E	QA0374
FINITE MARKOV CHAINS	59	KEMENY JOHN G	QA0273
FIXED NITROGEN	32	CURTIS HARRY ALFRED	TP0245.N8
FLOTATION ED. 2	57	GAUDIN ANTOINE M	TN0523
FLOW MEASUREMENT & CONTROL	59	COXON WILFRED F	TC0177
FLOW MEASUREMENT WITH ORIFICE METERS	51	STEARNS REID F	TC0177
(FLOW METES ENGINEERING HANDBOOK ED.2	46	BROWN INSTR CO	TC0177)
FLOW OF GASES THROUGH POROUS MEDIA	56	CARMAN PHILIP C	QC0168
FLOW OF HEAT IN METALS	42	AUSTIN JAMES B	QC032
FLOW OF NATURAL GAS	56	SMITH ROBERT V	
FLOW OF RAREFIED GASES	61	SCHAAF	
FLUID AGITATION HANDBOOK	56		
FLUID & PARTICLE MECHANICS			
FLUID DYNAMICS		60	TA0161
FLUID DYNAMICS		58	QA0935
FLUID DYNAMICS & HEAT TRA			TP0986
FLUID DYNAMIC		59	QC0189
FLUID FLO		58	QC0189
FLUID		43	QD0095

MASTER CATALOG ARRANGED BY AUTHOR

	. . . ED. 2	45	QD0022
WHITNEY	LES & PROBLEMS IN ENERGETICS	55	QC0073
	HEAT INERTIA IN PROB OF AUTO CONTROL OF TEMP	50	TH7226
	ADVANCED MATH IN PHYSICS & ENGR	53	QA0037
. POLY INST	PROCFEDINGS OF SYM ON INFORMATION NETWORKS	55	TK5101
BROOKLYN POLY INST	PROCFEDINCS OF SYM MOD ADV IN MICROWAVE TECH	55	TK7801
BROOKLYN POLY INST	PROCEEDINGS OF SYMPOSIUM ON MOD NETWORK SYN	52	TK3226
BROOKLYN POLY INST	PROCEEDINGS OF SYMPOSIUM ON MOD NETWORK SYN	55	TK3226
BROOKLYN POLY INST	PROCEEDINGS OF SYM ROLE SOLID STATE PHENOMENA	57	QC0501
BROWN AUBREY I	INTRODUCTION TO HEAT TRANSFER ED. 2	51	QC0320
BROWN AUBREY I	INTRODUCTION TO HEAT TRANSFER ED. 3	58	QC0320
BROWN CLINTON CARL	ELECTRONIC INSTRUMENTATION FOR BEHAVIORAL SCI	58	QP0357
BROWN GEORGE G	NATURAL GASOLINE & VOLATILE HYDROCARBONS	48	TP0692.2
BROWN GEORGE G	UNIT OPERATIONS	50	TP0155
BROWN RORERT GROVER	LINES, WAVES* & ANTENNAS	61	TK3001
BROWN SANBORN C	BASIC DATA OF PLASMA PHYSICS	59	QC0711
BROWN THOMAS B	LLOYD W TAYLOR MAN ADV UNDER-GRAD EXPER PHYS	59	QC0033
(BROWN INSTR CO	FLOW METER ENGINEERING HANDBOOK ED.2	46	TC0177)
BROWNE CHARLES A	HISTORY OF THE AMER CHEM SOC	52	QD0002
BROWNELL LLOYD E	PROCESS EQUIPMENT DESIGN	59	TP0157

FIG. 2-8 Portion of a short-title printed book catalog.

classification number. Figure 2-9 is the same type of book catalog in use at another library, and is shown here to illustrate the differences in format and arrangement that are possible while still using the same classification scheme.[43]

The KWIC indexes, described earlier in this chapter, usually provide a one- or two-line title listing, with each title repeated in the list as many times as there are significant index terms in the title, and listed alphabetically by each of the subject terms (see Figs. 2-2 and 2-3). Additional index terms as well as "see" and "see also" references can be inserted into this display.[44]

Arranged by a Planned Scheme. Most subject indexes are alphabetical listings of subject headings,

[43] Griffin, M., "Printed Book Catalogs," *Revue de la Documentation,* Vol. 28, No. 1, pp. 8–17 (February 1961).

[44] Veilleux, M., cited earlier in this chapter, fn. 12.

AUTHOR		TITLE	V.	CALL NUMBER	DATE	LOC.
BRUIJN	NG	ASYMPTOTIC METHODS IN ANALYSIS		QA 221 B892	58	RE
BRUINSMA	AH	REMOTE CONTROL BY RADIO		TK 6570 B 89	52	RE
BRUINSMA	AH	PRACTICAL ROBOT CIRCUITS		TJ 211 B892	60	RE
BRUMMEL	L	UNION CATALOGUES		Z 695 83B893	56	RE
BRUNAUER	S	ADSORPTION OF GASES AND VAPORS	1	QC 182 B825	43	RE
BRUNER	JS	STUDY OF THINKING		BF 455 B 75	56	RE
BRUNER	JS	PROCESS OF EDUCATION		LB 885 B894	60	RE
→ BRUNING	H	SECONDARY ELECTRON EMISSION		PC 721 8 86	54	RE

TITLE	AUTHOR		V.	CALL NUMBER	DATE	LOC.
SCREW THREAD STANDARDS FOR FED SERVICES	U S ISTC		3	RTJ 1340 U005	57	RE
SCULPTURE OF THE WESTERN HEMISPHERE	I B M			NB 30 1610	42	RE
SEA AROUND US	CARSON	RL		6C 21 C 3	51	RE
SEALS BOOK	MACH DESIGN			RTJ 1040 M149	61	AS
SEARCH ING THE CHEM ICAL L ITERATURE	A C S			QD 1 A512	61	RE
SEARCHING THE LITERATURE FOR TRANSDUC...	PEARLSTEIN	J		Z 5834 P359	60	RE
SECOND COLOR	COYNE	RS		Z 258 C881	61	RE
SECOND ENCYCLOPEDIA OF STORIES QUOTATION	BRAUDE	JM		RPN 6081 B825	57	RE
SECOND SCIENTIFIC AMERICAN BOOK OF MA...	GARDNER	M		PA 95 6228	61	RE
→ SECONDARY ELECTRON EMISSION	BRUNING	H		QC 721 B 86	54	RE
SECRET AND URGENT	PRATT	F		Z 104 P914	42	

SUBJECT

CALL NO.		BOOK			DATE	LOC.
		COLLISIONS NUCLEAR PHYSICS				
		SEE ALSO RADIATION				
QC 173	A881	ATOM MOL BEAM		PHOC ATOMIC AND MOLECULAR BEAMS CO	60	RE
→ QC 721	B 86	BRUNING	H	SECONDARY ELECTRON EMISSION	54	RE
PC 721	I615	INTL CONF PHY		PROC INTL CONF ON THE PHYSICS	61	RE
PC 721	M416	MASSEY	HS	ELECTRONIC AND IONIC IMPACT PHENOM	52	RE
QC 721	O110	OAK RIDGE		ATOMIC AND MOLECULAR COLLISION CRO	61	RE
QC 721	R325	REED	RI	ION PRODUCTION BY ELECTRON IMPACT	62	RE
QC 771	U220	UEHLING	EA	PENETRATION OF CHARGED PART IN MAT	58	RE
		COLLOIDS				
		SEE ALSO AZEOTROPES				
		SEE ALSO RHEOLOGY				
QD 549	A 32	ALEXANDER	J	COLLOID CHEMISTRY	50	RE
QD 172	A04	F254 FAIT	I	ALKALI MtIAL DISPERSIONS	61	
QD 549	I004	ILER	RK	COLLOID CHEMISTRY OF SILICA + SILI		
QD 549	J610	JIRGENSONS	B	ORGANIC COLLOIDS		
QD 549	K940	KRUYT	IHR	COLLOID SCIENC•		
QD 549	K940	KRUYT				
QD 549						

FIG. 2-9 Portion of a short-title printed book catalog.

under which some reference to pertinent file items appears. The index to this book is a good example of such an index. The simplest form of this type of index is one which uses a few broad subject headings in alphabetic order and lists under each of these headings the numbers of the pertinent file items (see Fig. 2-10). A classification scheme, rather than an alphabetic array of terms is another method of display. One example of such a classified subject index is given in Fig. 2-11.

A more advanced subject index is the index in which the indexing terms are "co-related" in the display.[45] The indexing terms are, taken from a relatively small and well-controlled vocabulary, and are

[45] Bernier, C. L., et al., "Correlative Indexes. VII. Trope Vocabularies and Trope Indexes for Chemistry," *Journal* of Chemical Documentation, Vol. 2, No. 2, pp. 93–102 (April 1962).

Bernier, C. L., "Correlative Indexes. I. Alphabetical Correlative Indexes," American Documentation, Vol. 7, No. 4, pp. 283–288 (October 1956).

Subject Index to Applied Mechanics Reviews

Volume 14, January-December 1961

(Numbers used are serial numbers of reviews)

THE GROWTH OF THE FIELD AND THE NUMBER OF PAPERS PUBLISHED AND THEN REVIEWED IN AMR MADE MACHINE PROCESSING OF THE 1961 SUBJECT INDEXING MANDATORY. THE TABULATING EQUIPMENT USED FOR PRINTING THE INDEX HAS ONLY UPPER CASE LETTERS. FOR THIS REASON AN ASTERISK IS USED BEFORE AND AFTER THE SEE AND SEE ALSO WORDINGS. THE MAIN AND SUB-HEADINGS ARE DIFFERENTIATED BY INDENTATION. THE SUBHEADINGS ARE PRINTED IN PARENTHESIS IF THEY FOLLOW THE MAIN HEADINGS IN THE SEE ALSO GROUP SUCH AS
SEE ALSO STREAM (NATURAL)

A

ABLATION
433 434 435 439 998 1611 1612 2102 3777
3802 5604 5605 6305 6874

ABSORPTION OF SOUND--- *SEE* SOUND (ABSORPTION)

ACCELERATION MEASUREMENT
6682
SEE ALSO--- STRAIN CAGES

ACOUSTIC PROPERTIES OF MATERIALS
282 531 533 2252 2442 5096 6027 6835

ACOUSTICS
12 529 531 532 533 725 1100 1101 1102
1104 1105 1106 1107 1108 1109 1110 1693 2252
2257 2258 2760 3318 3413 3896 5099 5101 2200
2257 2258 2760 3318 3413 3896 5099 5101 5200
5738 5742 5743 5754 6375 6382 6383
SEE ALSO--- COMBUSTION (ACOUSTIC) -- INTERFACE PHENOMENA -- ULTRA AND INFRA-SONICS

ADHESIVES
FEBRUARY FEATURE ARTICLE

ADSORPTION
1629 3774

AERODYNAMIC COEFFICIENTS
347 348 1462 1467 1997 3164 3167 5504 6839
6931 6953
DRAG COEFFICIENT
3671 3681 3727 3728 3729 3863 3864 4286
4353 4354 4482 4485 4863 4874 4915 4918
5450 5461 5465 5503 6203 6756
LIFT COEFFICIENT
352 2330 2615 3730 3733 3734 3853 3863
3864 4286 4354 4482 4856 4915 4918 5450
5461 5465 5502 6160 6163 6164 6203 6742
6756 6795
MOMENT COEFFICIENT
349 354 1989 2330 2615 3734 3853 3864
4286 4292 4354 4482 4485 4856 5450 5461
5465 5502 6160 6163 6164 6203 6756 6795

AERODYNAMIC HEATING
311 510 654 1084 1954 1993 2066 2150 3100
3112 3286 3672 3673 3676 3678 3679 3680 3761
3863 4999 5183 6872
SEE ALSO---HEAT TRANSFER

AERODYNAMIC INTERFERENCE
350 351 908 1450 1452 1453 1465 1522 1568
3113 3166 3671 3730 3734 3853 4288 4293 4868
4875 4954 5457 5458 5459 5460 5505 5554 6159
6190 6206 APRIL FEATURE ARTICLE

AEROELASTICITY
503 507 508 1661 2224 2734 3285 3287 5070
5071 5073 5697 5698 5699 5700 5701 5702 5824
6163 6164 6348 6349 6350 6351 6352 6929 6930
6932 6933 6934 6935 6936 6937
SEE ALSO--- AIRFOIL AND WING, OSCILLATING-- FLUTTER -- GUST LOADS -- HYDROELASTICITY -- WINGS

AERONAUTICS
12 1069 1665 2736 2829 2830 3154 3167 3852
3853 3854 3855 3856 5706 5707 5708 5709 5710
5711 JULY FEATURE ARTICLE

AEROTHERMOACOUSTICS
1524

AEROTHERMOCHEMISTRY
305 309 338 339 2001 2569 2573 3681 3683
3685 3702 3716 3717 3762 3763 3808 4853 4901
4907 5489 5491 5495 5569

AEROTHERMODYNAMICS
331 455 952 990 991 1474 1613 1621 1998
2004 3102 3117 3119 3123 3126 3672 3673 3678
3679 3681 3761 3762 3763 3783 3813 5648 6848
APRIL FEATURE ARTICLE

SEE ALSO--- AERODYNAMIC HEATING -- HEAT TRANSFER (BOUNDARY LAYER) -- TEMPERATURE RECOVERY

AGING
235 236 776 777 1263 1819 2451 2452

AIRCRAFT STRUCTURES--- *SEE* STRUCTURES (AIRPLANE)

AIRFOIL CASCADES--- *SEE* CASCADES OF AIRFOILS

AIRFOIL CHARACTERISTICS
895 3158 4292 4913 5500 6789
PRESSURE DISTRIBUTION
349 931 1461 4289 4875 4918 4936 5457
5501 6153 6158 6747 6835
SEE ALSO--- AERODYNAMIC COEFFICIENTS,-- STALLING

Review Numbers:	1-582	583-1169	1170-1724	1725-2298	2299-2830	2831-3397	3398-3950	3951-4566	4567-5169	5170-5806	5807-6448	6449-7040
Issue:	January	February	March	April	May	June	July	August	September	October	November	December

1

FIG. 2-10 Portion of a conventional alphabetic subject index. (From Applied Mechanics Reviews.)

Influence of different acoustical stimuli on the threshold of the contralateral ear. M. Loeb and J. L. Fletcher. **33: 857—1961**

Measurement of the displacements and nonlinearities of the guinea pig tympanum. Lothar O. Hoeft, Eugene Ackerman, and Adam Anthony. **33: 1671—1961**

Minimum phase responses for the basilar membrane. James L. Flanagan and Carol M. Bird. **33: 830-1961**

Monaural double click interactions at the round window, eighth nerve, and inferior colliculus of cat. Alfred Finck and Robert J. Ruben. **33: 1671—1%1**

Present status of the neurophysiology of the cochlea. Hallowell Davis. **33: 1662(T)—1961**

Properties of the 8th nerve. Robert J. Ruben, Hugo Fish, and William Hudson. **33: 830—1961**

Quantitative evaluation of inertial and compressional components of bone conduction in cats. Juergen Tonndorf, R A. Campbell, L. Bernstein, and J. P. Reneau. **33: 1671—1961**

Shifts in the masked threshold. Ira J. Hirsh. **33: 859—1%1**

Steady'vs intermittent noise bursts for evoking and maintaining middle ear muscle reflex reactions. E. S. Mendelson and D. L. Kemmerer. **33: 858—1961**

SUBTITLES

Microphonics in animal ears. Edward A. Rice and Edward W. Shinabarger. **33: 923—1961**

4.3 Binaural Hearing. Localization

PAPERS

Auditory threshold shifts produced by simultaneously pulsed contralateral stimuli. Carl E. Sherrick, Jr., and Pedro L. Mangabeira-Albernaz. **33: 1381—1%1**

Binaural interaction of clicks of different frequency content. Bruce H. Deatherage. **33: 139-1961**

Lateralization vs localization. Lloyd A. Jeffress and Robert W. Taylor. **33: 482-1961**

Phasor analysis of some stereophonic phenomena. Benjamin B. Bauer. **33: 1536—1961**

Time vs intensity in the localization of tones. Randolph H. Whitworth and Lloyd A. Jeffress. **33: 925—1961**

LETTERS TO THE EDITOR

Binaural hearing aids. An enigma. James Jerger and Donald Dirks. **33: 537—1%1**

ABSTRACTS

Binaural detection of single frequency signals in the presence of noise. Mark B. Gardner. **33: 855—1961**

Binaural lateralization of cophasic and antiphasic clicks. James L. Flanagan. E. E. David, Jr.. and B. J. Watson. **33: 840—1961**

Head-position identification. H. N. Wright. **33: 855—1961**

Phasor analysis of the stereophonic phenomena. Benjamin B. Bauer. **33: 852-1961**

Study of intelligibility improvement with out-of-phase connections of headphones. D. Greenberg and H. O. Benecke. **33: 851—1961**

4.4 Deafness. (See **also** 5.5.)

PAPERS

Noise-induced permanent threshold shift at 2000 cps and 4000 cps. J. C. Nixon and A. Glorig. **33: 904—1%1**

ABSTRACTS

Cochlear potentials in the human being. Robert J. Ruben, John E. Bordley, and Alfred T. Lieberman. **33: 830—1961**

Equivalent circuit study of fenestration. G. F. Paskusz. **33: 831—1961**

Experimental case study of a "tone-deaf" individual. Wilson P. Tanner, Jr., and Laurence Rivette. **33: 1655—1961**

Recent research on noise—induced hearing loss. Aram Glorig. **33: 1662(T)—1961**

Temporary depression in VIIIth nerve output following 90-110 SPL in the nonanesthetized cat. J. Donald Harris. **33: 1670-1%1**

BOOK REVIEWS

Hearing and Deafness. H. Davis and S. R. Silvcrman, Editors. (Reviewed by James Jerger.). **33: 544—1961**

4.4.5 Auditory Fatigue. Temporary Threshold Shift

PAPERS

Contralateral threshold shift and reduction in temporary threshold shift as indices of acoustic reflex action. M. Loeb and J. L. Fletcher. **33: 1558—1961**

Effect of matching time on perstimulatory adaptation. Arnold M. Small, Jr., and Fred D. Minifie. **33: 1028—1961**

Exploratory studies on temporary threshold shift from impulses. W. Dixon Ward, Weldon Selters, and Aram Glorig. **33: 781—1%1**

Temporary threshold shifts from tones and noise bands of equivalent rms sound-pressure level. Paul O. Thompson and Robert S. Gales. **33: 1593—1961**

ABSTRACTS

Acoustic trauma: concomitant behaviotal and electrophysiological measurements in the cat. Dana Beatty and F. Blair Simmons. **33: 830—1961**

Contralateral threshold shift and reduction in temporary threshold shift as indices of acoustic reflex action. J. L. Fletcher and M. Loeb. **33: 1670—1961**

Damage risk criterion for and auditory fatigue from the Audiac. K. D. Kryter, A. Z. Weisz, and F. M. Wiener. **33: 858—1961**

Effect of auditory fatigue on differential intensity thresholds. Donald N. Elliott, Winifred Riach, and Herman Silbiger. **33: 859-1%1**

Effect of matching time on perstimulatory auditory adaptation. Arnold M. Small, Jr., and Fred D. Minifie. **33: 859—1961**

Poststimulatory fatigue of masked threshold. Robert C. Bilger and Charles V. Anderson. **33: 1656—1961**

Poststimulatory shift of the Preyer reflex threshold in rats. P. N. Herman, T. D. Clack, and J. D. Harris. **33: 1670—1961**

Shifts in the masked threshold. Ira J. Hirsh. **33: 859—1961**

Short-time aftereffects of noise on auditory nerve responses. W. T. Peake and J. F. Buoncristiani. **33: 1670—1961**

Temporary depression in VIIIth nerve output following 90-110 SPL in the nonanesthetized cat. J. Donald Harris. **33: 1670-1%1**

Temporary hearing losses for bare and protected ears as a function of exposure time to continuous and firing noise. Alexander Cohen. **33: 858—1961**

4.5 Instruments Relating to Hearing, and for the **Testing** of Hearing. (See also **5.5.)**

PAPERS

Improved method for studying tympanic reflexes in man. Emanuel S. Mendelson. **33: 146—1961**

LETTERS TO THE EDITOR

Binaural hearing aids. An enigma. James Jerger and Donald Dirks. **33: 537—1961**

Supra-aural cushions in audiometry. Russell L. Sergeant and J. Donald Harris. **33: 966—1961**

PATENTS

Acoustical device. John F. Rose. **33: 1645—1%1**

Apparatus for measuring psychological-physiological relationships. Monroe T. Hunsicker and Robert H. De Vaney. **33: 1438—1%1**

Binaural eyeglass hearing aid construction. Sam Posen and Louis A. McNabb. **33: 254—1%1**

FIG. 2-11 Portion of a conventional classified subject index. (From The Journal of the Acoustical Society of America.)

not simply words from the text. Coded representations of the subject headings for a specific file item are grouped together with the item description or number, and listed under each of the index term numbers, so that the user can see what other indexing terms were used with that file item. An example of a section of one such correlative index by Bernier is given in Fig. 2-12.[46]

One way to make an even more useful display is to provide, in addition to the grouping of references under prescribed subject headings, some small amount of more substantive detail about the contents of the file items. This descriptive information (a modifier line or micro-abstract) might consist of a single line or less which describes the text of the file item. One sample of such an index, taken from Nuclear Science Abstracts, a publication of the U.S. Atomic Energy Commission, is shown in Fig. 2-13. Another example in which the micro-abstract briefly describes the major finding or content of the paper is the Index Handbook of Cardiovascular *Agents*.[47]

Two other types of indexes which arrange and display the information in a special way in order to facilitate searching are O'Connor's Scan-Column Index and Ledley's Tabledex—however, they have not received any extensive use to date.[48]

OTHER TYPES OF INDEXES

In addition to the subject indexes described in the last section and the bibliographical indexes that include the descriptive indexing information and might be arranged in order by authors, corporate authors, report number, and contract number, there are also some other special types of indexes.

One of these is the citation index, which records and displays cited references, and by whom cited. This type of index (e.g., Shepard's Citations) has been used by the legal profession for many years to trace the citations of the courts and has developed a very compact method of notation and display (see Fig. 2-14).[49] There has been recent interest in applying the same technique to the indexing of citations to technical articles given in the scientific journals (see illustrations in Chapt. 7). This index would, in fact, be an index to authors in footnote references or bibliographies, and might be of value in supplementing existing reference tools.[50]

Another example of a special type of index is the index of chemical compounds as typified by Index Chemicus. In this index the chemical elements of a compound reported in the literature are permuted and listed in alphabetical order with the context of the accompanying elements in the compound, to make it relatively easy to locate specific compounds and a reference to the literature that mentioned them.

CRITERIA FOR THE SELECTION OF AN INDEXING SYSTEM

There is of course no best indexing system for all jobs, since each job has different requirements in terms of such things as size of file, search complexity, staff, and funding for the support of the system. In general, the following points should be considered during the review of an existing indexing method or during selection of a new system:

1. Type of ultimate user (the users vary in needs, habits, and approaches)

[46] A sample of a trope index of ***Chemical Abstracts*** is given in Bernier's *J.* ***Chem. Doc.,*** article cited above, fn. 45.

[47] Welt, I. D., ***Index Handbook of Cardiovascular Agents: 1961–1966*** (National Academy of Sciences—National Research Council, Washington, D.C., 1960).

MacMillan, J. T., and I. D. Welt, "A Study of Indexing Procedures in a Limited Area of the Medical Sciences," ***American Documentation,*** Vol. 12, No. 1, pp. 27-31 (January 1961).

[48] O'Connor, J., "The Scan-Column Index," ***American Documentation,*** Vol. 13, No. 2, pp. 204–209 (April 1962).

Ledley, R. S., "Tabledex: A New Coordinate Indexing Method for Bound Book Form Bibliographies," ***Proceedings of the International Conference on Scientific Information,*** Vol. 2, pp. 1221–1243 (National Academy of Sciences, Washington, D.C., 1959).

Ledley, R. S., et al., ***Selected Bibliography of the International Geophysical Year: An Example of Tabledex Formats,*** Report 62071/18100 of the National Biomedical Research Foundation (undated).

[49] ***Shepard's Federal Labor Law Citations*** (Shepard's Citations, Inc., Colorado Springs, Colorado). A compilation of citations to Decisions and Orders of the National Labor Relations Board, United States Supreme Court Decisions in Labor Cases, Lower Federal Court Decisions in Labor Cases, and Labor Provisions in the United States Code.

[50] Lipetz, B., "Compilation of an Experimental Citation Index from Scientific Literature," ***American Documentation,*** Vol. 13, No. 3, pp. 251–266 (July 1962).

Garfield, E., "Citation Indexes for Science: ***Science,*** Vol. 122, No. 3159, pp. 108–111 (July 15, 1955).

Tukey, J. W., "Keeping Research in Contact with the Literature: Citation Indexes and Beyond," ***Journal of Chemical Documentation,*** Vol. 2, No. 1, pp. 34-37 (January 1962).

11 92	HEAT TRANSFER TO WATER IN AN ANNULUS.	2214H
11 94	EFFICIENT FURNACES FOR MAGNETIC ROASTING OF ORES.	2385G
11 95	THE ENTHALPY AND HEAT CAPACITY OF MAGNESIUM ANO OF TYPE-430 STAINLESS STEEL FROM 700 TO 1100-DEG.-K.	2267B
11 95	EXPERIMENTS IN THE BASIC CUPOLA FURNACE WITH HOT BLAST AND COKE INCRUSTATION.	2386F
11 95	BACKGROUND FOR PRACTICAL HEAT-TREATMENT OF VARIOUS TITANIUM ALLOY TYPES.	2391D
11 95	HEAT-TREATMENT OF HIGH-STRENGTH NODULIZED CAST IRON.	2391G
11 96	ELIMINATION OF CARBON IN THE OPEN-HEARTH FURNACE. I.	2386E
11 96	INFLUENCE OF ELECTRIC HEATING ON PHASE TRANS-FORMATION IN PLAIN CARBON AND IN CHROMIUM STEELS.	2394C
11 96	EFFECT OF PRETREATMENT OF MARTENSITE ON SUB-SEQUENT GRAPHITIZATION AT 1200-OEGS.-F.	2394I
11 96	HEAT-TREATMENT OF HAROENABLE COPPER-NICKEL-SILICON ALLOYS.	2402A
11 97	THE MELTING OF MULLITE.	2938G
11 97	POSSIBILITY OF RAPID BURNING CYCLE IN THE CERAMIC INDUSTRY.	2939D
11 97	MEASUREMENT OF THERMAL CONDUCTIVITY OF REFRACTORIES AT HIGH TEMPERATURES.	2942D
11 98	APPARATUS FOR PREHEATING AND ROASTING THE CHARGE IN CEMENT KILNS.	2947C

12 REACTIONS

12	CONTINUOUS STIRREO TANK REACTORS.* A NEW GRAPHICAL METHOO FOR COMPLEX REACTIONS AND REFLUX DESIGNS.	2215E
12	CHEMICAL ENGINEERING FUNDAMENTALS. WHY KINETICS FOR REACTOR DESIGNS.	2216A
12	GENERAL CONSIDERATIONS IN REACTOR DESIGN.	2216A
12	REACTOR OESIGN PROBLEM.* SIMPLE REACTION.	2216A
12	CHEMICAL REACTOR STABILITY AND SENSITIVITY.	2216G
12	CHEMICAL PROCESSING AT HIGH TEMPERATURE ANO HIGH PRESSURE.	2217D
12	THEORY OF PH.	2219C
12	THE MODERN STATE OF THE PROBLEM OF CATALYSIS ANO THE THEORETICAL BASES FOR THE STUOY OF CATALYSTS.	2258G
12	SELECTION OF MATERIAL FOR CONTACT CATALYSTS.	2894B
12 2 9	THE USE OF HOMOPOLAR ANO HETEROPOLAR CHROMATOGRAPHY IN MEDICINE.	2719B
12 2 48	VEGETABLE TANNING. XI. TWO NEW METHODS FOR OETERMINING THE ACIDITY OF VEGETABLE-TANNEO LEATHER.	2997I
12 3	AN EVALUATION OF THE ACCURACY OF THE METHOD OF STATIONARY CONCENTRATIONS (BODENSTEIN/S METHOD).	2254G
12 3 7 24	KINETICS OF THE DISSOCIATION OF WEAK ACIDS AND BASES. APPLICATION OF POLAROGRAPHY AND VOLTAMMETRY AT CONSTANT CURRENT.	2251C
12 3 7 95	THE MICROELECTROCHEMICAL METHOD OF INVESTIGATION OF STRUCTURAL CORROSION.	2318I
12 3 26	A METHOO FOR THE DETERMINATION OF THE SURFACE-VOLUME MECHANISM OF CATALYTIC REACTIONS.	2261H
12 3 45	CORRELATION OF ACCESSIBILITY OF CELLULOSE WITH ACETYLATION RATE.	2969G
12 3 45	ALCOHOLYSIS OF CELLULOSE.	2970G
12 3 84	THE OXIOATION STABILITY OF LUBRICATING OILS. ITS SIGNIFICANCE AND DETERMINATION.	2964H
12 3 88	A CONTINUOUS CONDUCTOMETRIC METHOD FOR MEASURING THE RATE OF HYDROLYSIS OF CHLORIOES.	2256H
12 4 5 13 36	INFLUENCE OF THE CONCENTRATION ON THE OPTICAL PROPERTIES OF SOLUTIONS OF 3,6-DIAMINOACRIOINE.	2291C
12 4 6 88	THE ACTIVITY COEFFICIENT OF HYOROCHLORIC ACIO IN CONCENTRATED AQUEOUS HIGHER-VALENCE-TYPE CHLORIDE SOLUTIONS AT 25-OEG. III. THE SYSTEM HYDROCHLORIC ACID-ALUMINUM CHLORIDE.	2251B
12 4 7	ELECTROCHEMICAL PROPERTIES OF AN ION-EXCHANGE MEMBRANE.	2330H
12 4 9 33 88	THE EXTRACTION OF METAL COMPLEXES. XX. THE DISSOCIATION CONSTANTS AND PARTITION COEFFICIENTS OF 1-NITROSO-2-NAPHTHOL AND 2-NITROSO-1-NAPHTHOL.	2252B
12 4 23 24	THE POTENTIOMETRIC MEASUREMENT OF ACID DISSOCIATION CONSTANTS AND PH IN THE SYSTEM METHANOL-WATER. PKA VALUES FOR CARBOXYLIC ACIDS AND ANILINIUM IONS.	2252I
12 4 24	EFFECT OF STRUCTURAL FACTORS ON CATALY7IC DECOMPOSITION OF ALCOHOLS OF DIFFERENT MOLECULAR WEIGHTS.	2263D
12 4 26	THE INFLUENCE OF THE SOLVENT ON REACTION VELOCITY. XI. ALKALINE HYDROLYSIS OF ETHYL BROMIDE IN ACETONE-WATER MIXTURES.	2260E
12 4 29	HYDROGEN FLUORIDE AS A CONDENSING AGENT. IV. HYDROGEN FLUORIDE AS A SOLVENT FOR CATALYTIC REDUCTION.	2507G
12 4 84	TRANSFER-LINE REACTOR AND CYCLONE SEPARATOR.	2970E
12 4 88	THE DETERMINATION OF THE BISULFATE DISSOCIATION QUOTIENT FROM POTENTIOMETRIC MEASUREMENTS.	2251F
12 4 88	ACTIVITY COEFFICIENTS IN 7HE SYS7EMS HC1-(LI, NA, K, NH4)CI-H2O.	2251I
12 4 88	OXYGEN EXCHANGE IN THE HYDROTHERMAL GROWTH OF QUARTZ.	2254E
12 5 9	CHARACTERIZATION OF PHYSICAL ADSORPTION SYSTEMS. III. THE SEPARATE EFFECTS OF PORE SIZE AND SURFACE ACIDITY ON THE ADSORBENT CAPACITIES OF ACTIVATED CARBONS.	2239A
12 5 11 13 41	PHOTOTROPIC BEHAVIOR OF 4-(P-DIMETHYLAMINOPHENYLAZO)- PHENYLMERCURIC ACETATE.	2303B
12 5 24	RETENTION OF CONFIGURATION IN THE WOLFF REARRANGEMENT.	2629A
12 5 29 87	STERIC EFFECTS IN DISPLACEMENT REACTIONS. II. THE RATES OF REACTION OF ALKYL IOOIDES WITH THE MONOALKYLPYRIDINES. STERIC STRAIN IN THE ACTIVATED COMPLEX.	2585F
12 5 30	THE EFFECT OF CONFIGURATION ON THE REACTIVITY OF THE CHALCONE SYSTEM.	2477A
12 5 30	OXIDATION OF HINDERED PHENOLS. III. 7HE REARRANGEMEN7 OF THE 2,6-DI-TERT-BUTYL-4-METHYLPHENOXY RADICAL.	2514G
12 5 31	1,3-SHIFTS. III. THE KINETICS OF THE THERMAL REARRANGEMENT OF PHENYL BENZANILIMINO ETHERS.	2464E
12 5 36	OZONOLYSIS OF QUINOLINE, 6,7-DIMETHYLQUINOLINE, 5,8-DI-METHYLQUINOLINE, AND 2,3-DIMETHYL-QUINOLINE IN CONNECTION WITH THE REACTIVITY OF THE BONDS IN THE RING SYSTEM.	2595D
12 5 37	SPECTROPHOTOMETRIC STUDY OF THE HYDROLYSIS CONSTANTS OF THE NEGATIVE IONS OF SOME ARYL IMIDAZOLES.	2253D
12 5 43	ASYMMETRIC SYNTHESES BY THE GRIGNARD REACTION.	2500E
12 5 44	INFLUENCE OF STERIC HINORANCE ON REACTIVITY OF ARYLAZOGLUCAMINES.	2460H
12 5 48	INFLUENCE OF SUBSTITUENTS ON CHEMICAL REACTIVITY. VIII. NEW A-AMINO ACIDS FORMED BY THE INTER-ACTION OF AZLACTONES WITH GRIGNARO COMPOUNDS.	2498E
12 5 88	STRUCTURE OF CHROMIUM OXIDE AND THE MECHANISM OF ITS HIGH-TEMPERATURE OXIOATION.	2230D
12 5 88	CHAIN BRANCHING IN GLASSY POLYPHOSPHATES.* DEPENOENCE ON THE SODIUM-PHOSPHORUS RATIO ANO RATE OF DEGRAOATION AT 25-DEG.	2245B
12 7	SOME PROBLEMS OF THE ELECTRON THEORY OF CATALYSIS BY SEMICONDUCTORS.	2261B
12 7	POLAROGRAPHIC REACTION KINETICS.	2515C
12 7	THE THEORY OF THE POLAROGRAPHIC CURVE. VI. REACTIONKINETICALLY DETERMINEO POLAROGRAPHIC CURRENT DENSITY. 3. CHEMICAL REACTIONS FOLLOWING THE ELECTRODE PROCESS AND REGENERATING THE DEPOLARIZING AGENT.	2315D
12 7	OUR CONCEPTS CONCERNING THE KINETICS OF ELECTROOE PROCESSES.	2326F
12 7	METHOOS OF INVESTIGATING THE KINETICS OF ELECTRODE PROCESSES.	2326I
12 7	DIFFERENTIATION OF DIFFERENT ELECTROOE MECHANISMS BY MEASUREMENT OF TRANSITION TIME AT CONSTANT CURRENT IN STATIONARY SOLUTION.	2329G
12 7 9	ABSORPTION PHENOMENA AND ELECTROCHEMICAL KINETICS.	2325I
12 7 18	CHEMICAL EFFECTS FROM EXCITATION BY SECONDARY ELECTRONS PRODUCED BY IONIZATIONS RAYS.	2303D
12 7 24	LIMITING CURRENTS IN THE REDUCTION OF FORMALDEHYDE AT THE DROPPING MERCURY CATHODE.	2317A
12 7 81	ELECTROLYTIC CONDENSERS.	2331A
12 7 87	POLAROGRAPHIC STUDY OF TRIS(ETHYLENEDIAMINE)-CO(III) ION.	2316E
12 7 88	EFFECTS OF OXYGEN ON LEAD SULFIOE FILMS.	2223H
12 7 88	KINETICS OF CATHODIC HYDROGEN EVOLUTION.	2327B
12 7 88 95	REACTIONS OF METAL IONS AT MERCURY ELECTROOES IN FUSED SALTS.	2327E
12 7 95	ANODIC POLARIZATION OF ZIRCONIUM AT LOW POTENTIALS. FORMATION RATES, FORMATION FIELD, ELECTROLYTIC PARAMETERS, ANO FILM THICKNESSES OF VERY THIN OXIDE FILMS.	2319B
12 7 95	ELECTRODE POTENTIAL AND CORROSION OF METALLIC MANGANESE IN AMMONIUM CHLORIDE SOLUTION.	2319G
12 7 95	CAUSES OF CORROSION INHIBITION OF ZINC CATHODES IN THE PRESENCE OF SURFACE-ACTIVE AOOI7IVES.	2320E
12 7 95	KINETICS OF ELECTRODE PROCESSES. III. EFFECT OF AODITIVES UPON THE RATE OF CONTACT OEPOSITION OF COPPER ON IRON.	2325H
12 7 95	FINAL REPORT ON FOUR ANNUAL ANODE INSPECTIONS. A REPORT OF THE TECHNICAL UNIT COMMITTEE T-28 ON ANODES FOR IMPRESSED CURRENT.	2406B
12 7 95	INDUCED ALTERNATING CURRENT USED FOR CATHOOIC PROTECTION OF A COATED PIPE LINE.	2406C
12 7 95	INHIBITION OF METALLIC CORROSION IN AOUEOUS MEOIA. A REVIEW.	2406G
12 7 95	EFFECT OF THE ADDITION OF AMMONIUM CHROMATE ON THE POTENTIAL ANO THE CORROSION OF MANGANESE IN AMMONIUM CHLORIDE.	2407A
12 7 96	EFFECT OF COMPOSITION ON THE CORROSION POTENTIAL OF IRON ALLOYS.	2406F
12 9	CONTINUOUS REACTIONS IN DISTILLATION EQUIPMENT.	2211A
12 9	GAS ABSORPTION ACCOMPANIEO BY CHEMICAL REACTION.	2212H
12 9	FEEOING GRANULAR PARTICLES INTO A GAS LIFT.	2218D
12 9	ADSORPTION AND SILVER-SILVER ION EXCHANGE.	2240D
12 9 11	SILICONE-COATED CATALYST.	2895A
12 9 44	CHROMATOGRAPHIC ADSORPTION. IV. CATION-EXCHANGE RESINS AS CATALYSTS IN GLYCOSIDE FORMATION.	2435D
12 9 84	LIGHT-ENOS RECOVERY IN FLUID HYOROFORMING.	2967A
12 9 93	CATALYTIC FUME INCINERATION.	2218E
12 10 11 85	FLOW EXPERIMENTS IN STUDYING KINETICS. ROLE OF DIFFUSION AND INFLUENCE OF FLOW PATTERN IN GASIFICATION REACTIONS.	2953H
12 10 11 88	THE ADIABATIC EXPONENT FOR CHEMICALLY CHANGING GAS MIX7URES.	2227C
12 11	EOUIVALENT ISOTHERMAL TEMPERATURES FOR NONISOTHERMAL REACTORS.	2215D
12 11 24	ACID CATALYSIS OF THE HYOROLYSIS OF ACE7IC ANHYDRIDE IN ACETIC ACID.	2260A
12 11 26 56	GASEOUS DETONATIONS. VII. A STUDY OF THERMODYNAMIC EQUILIBRATION IN ACETYLENE-OXYGEN WAVES.	2255F
12 11 41	THE OISSOCIATION OF SODIUM ACETYLIDE.	2254F
12 11 50	KINETICS OF HEAT OF INACTIVATION OF RIBONUCLEASE.	2699B
12 11 80	THE INTERNAL STRUCTURE OF FLAME REACTION ZONES.	2947I
12 11 84	CATALYTIC FLUID HYDROFORMING WITH INERT HEAT-TRANSFER SOLIDS.	2967C
12 11 85 88	IGNITION TEMPERATURES OF LEAD COMPOUND-CARBON MIXTURES.	2962G
12 11 88	THE THERMODYNAMIC STABILITY OF REFRACTORY BORIDES.	2254D
12 11 88	EFFECT OF TEMPERATURE ON THE RATE OF FORMATION OF NICKELIFEROUS ANTIGORITE.	2258C
12 11 88	THERMODYNAMICS AND KINETICS OF THE NITROGEN DIOXIDECATALYZED OXIDATION OF NITROSYL CHLORIDE TO NI7RYL CHLORIOE.	2259D
12 11 95	PREVENTING SCALING ON HEATING.	2412D
12 11 95	EVALUATION OF FUELS AND OXIDANTS FOR WELDING AND ASSOCIATED PROCESSES.	2947H
12 13	FLUORESCENCE AND AVERAGE LIFETIME OF EXCITED OH IN FLAMES.	2501E
12 13 24	KINETICS OF THE THERMAL AND PHOTOCHEMICAL OXIDATION OF ACETALDEHYDE.	2256C
12 13 24	PHOTOSENSITIZED DECOMPOSITION OF ACETYL PEROXIDE.	2302I
12 13 41	REVERSIBLE FLASH BLEACHING OF CHLOROPHYLL.	2302B
12 13 43	TRANSFORMATIONS OF ORGANIC COMPOUNDS CAUSEO BY ALUMINOSILICATE CATALYSTS. VIII. CHANGE IN THE OPTICAL ACTIVITY OF LIMONENE.	2499I
12 15 36	QUATERNIZATION KINETICS. I. SOME PYRIDINE DERIVATIVES IN TETRAMETHYLENE SULFONE.	2257G
12 15 88	SPECTROPHOTOMETRIC S7UDY OF THE KINETICS OF FERRIC THIOSULFATE REACTIONS.	2339G
12 19	PROBLEM OF THE RELATIONS BETWEEN RADIOCHEMICAL YIELD AND THE PHYSICAL ASPECTS OF THE QUALITY OF	2314G

FIG. 2-12 Portion of a correlative index.

creep behavior, effects of heat treatment, 16: 552
MAGNETIC FIELDS
charged-particle behavior in plasma-free, 16: 1115 (NYO-9495)
charged particle orbit in, 16: 607
measurement, performance of Hall probes, 16: 328 (BNL-5750)
micropulsations in aurðral zone, relations of enhancement with electron bremsstrahlung, 16: 920
particle trajectory in, recordings by spark chambers, 16: 374
plasma compression by, theory of cylindrical, 16: 1113 (NP-10852)
plasma containment by external and internal, 16: 1153
plasma dynamics in, 16: 1118 (SIT-P30)
plasma ring motion in coaxial, 16: 1274 (NASA-TR-R-121)
stability effects on conducting fluid sphere, 16: 881
stabilization, in beta spectrometer, 16: 357
MAGNETIC MIRROR MACHINES
plasma injection, 16: 1151
MAGNETOHYDRODYNAMIC CELLS
viscosity effects, 16: 1126
MAGNETOHYDRODYNAMICS
combustion in, 16: 865
detonation and ionization shock waves, 16: 1149
effects of magnetic field on gas flow in horizontal tube with heat difference, 16: 1129
equations for plasma containment, 16: 618
linear oscillations in relativistic plasma, 16: 1108 (ARL-TR-60-274 (Pt.II))
mechanics, 16: 603
of plasma, 16: 1118 (SIT-P30)
plasma stability conditions in, 16: 1114 (NP-10876)
radial flow of an electroconducting gas in a moving magnetic field, 16: 1128
stability of axial current fluid, effect of surrounding neutral medium, 16: 602
stability of electrical conducting inviscid fluid in axial current, coaxial conducting cylinder, and longitudinal magnetic field, 16: 600
stability of electrical conducting inviscid fluid slab in homogeneous magnetic field, 16: 601
stability of plasma in static equilibrium, 16: 1111 (NP-10848)
streaming in cylindrical tube, exact solution, 16: 859
theory, generalization of Oseen's solutions, 16: 1116 (PLR-75)
wave phenomena in a magnetoplasma, 16: 1109 (BSR-455)
wave propagation in ionosphere, 16: 1103 (AFCRL-64)
MAIZE
metabolism of corn roots, effects of 2,4-D and sugars, 16: 28 (ORO-470)
MALONIC ACID
electron spin resonance of gamma irradiated, 16: 232
ethyl, electron spin resonance of free radicals in gamma irradiated, 16: 115
MALONIC ACID, ESTERS
electron spin resonance of x-ray irradiated single crystals of potassium hydrogen, 16: 234
MAMMARY GLAND
radiotherapy, effects on cardiac muscle, 16: 61
MANGANESE
determination by solvent extraction and spectrophotometry using 1-(2-pyridylazo)-2-naphthol, 16: 139
determination in ore by activation analysis, 16: 689
effects on corrosion of Zircaloy-3 by steam and water, 16: 486 (BM-RI-5865)
MANGANESE BORATES
polymorphism, x-ray analysis, 16: 162
MANGANESE BROMIDES
spectra of tetra, arrangement of bromide ions, 16: 92(R) (TID-13601)
MANGANESE CHLORIDES
spectra of tetra, arrangement of chloride ions, 16: 92(R) (TID-13601)
MANGANESE FILMS
properties as neutron-measuring probes, 16: 336 (NP-10851)
MANGANESE IONS
catalytic effects on reactions of hydrogen peroxide with uranium(VI) ions, 16: 111
MANGANESE ISOTOPES Mn-54
determination in Bruderheim chondrite, 16: 917
MANGANESE ISOTOPES Mn-55
nuclear resonance in antiferromagnetic potassium manganese fluorides, 16: 616
MANGANESE-RHENIUM ALLOYS
phase studies, 16: 559
MANNITOL
uptake in corn roots, 16: 28 (ORO-470)
MANNOSE
uptake in corn roots, 16: 28 (ORO-470)
MARCOULE REACTORS
stack-disposal fall-out patterns, normal and accidental, 16: 395 (CEA-1924)
MARITIME GAS COOLED REACTOR
containment and safety systems, review, 16: 1250(R) (NDA-2159-2)
MASS SPECTROMETERS
adaptation of Consolidated Model 21-611-1, in analysis of uranium for hydrogen, 16: 261(R) (NCW-1465)
design and operation of two-dimensional magnetic prism, 16: 376
design and performance of high resolving power, 16: 355(T) (NP-tr-791)
design for detection of free radicals in gases, 16: 338 (NP-10899)
MATERIALS TESTING
ultrasonic techniques for ceramic coatings, 16: 389(R) (ARF-6043-20)
MATHEMATICS
continuous weapon casualty functions, step-function approximation, 16: 3 (SC-2992(TR))
differentials and gradients of functionals, 16: 477 (T) (UCRL-Trans-170)
dimensional consistency of equations for calculation of maximum ductile rupture in pressurized systems, 16: 392 (WAPD-TM-56(Add.))
relations between cross section in the center of mass system and that in the laboratory system, 16: 1019 (AROD-2188:2)
theory of dimensional analysis and model testing, 16: 303(T) (AEC-tr-4850)
theory of operational equations, variational principles in, 16: 476(T) (UCRL-Trans-169)
use for determination of normalized response of two coupled mass spring systems, 16: 589 (SCR-408)
vorticity transport equation, solutions, 16: 317 (SC-4640(RR))
MATTER
energy of highly compressed, statistical theory, 16: 1163(T) (UCRL-Trans-722(L))
theory of highly compressed, 16: 1169
MECHANICAL ENGINEERING
analysis of the spike phase pulse in two-phase shock tests, 16: 390 (SCTM-316A-60(73))
drawing dimensioning, positional tolerance method for, 16: 300 (SCR-83A)
theory of dimensional analysis and model testing, 16: 303(T) (AEC-tr-4850)
MECHANICS
nonlinear systems analogy to classical, 16: 609
MEDICINE
research progress in, 16: 21(R) (ACRH-15)
MEMORY DEVICES
design and circuitry for 256-channel time analyzer, 16: 350(T) (AEC-tr-4700)
MERCURY
determination by solvent extraction and spectrophotometry using 1-(2-pyridylazo)-2-naphthol, 16: 139
hydrogen reactions at −195°C, 16: 112
MERCURY IODIDES
effects on aluminum oxidation rates, 16: 484(R) (AD-256309)
MERCURY IONS
catalytic effects on reactions of hydrogen peroxide with uranium(VI) ions, 16: 111
MERCURY ISOTOPES Hg-197
hyperfine structure of 24-hr isomer, 16: 1055
isomeric yields from (d,2n) reactions on gold-197, 16: 1025 (NP-9365)
MERCURY ISOTOPES Hg-201
nuclear magnetic moments, 16: 1054
MESONS
interactions of neutral vector, with photons, 16: 967

FIG. 2-13 Portion of a conventional alphabetic subject index with micro-abstracts.

UNITED STATES SUPREME COURT REPORTS (Labor Cases) Vol. 350

117Bd 440
No 60
j117Bd 443
No 60
81Az 168
302P2d 934
'57ALC278
31LC¶70286
39LRM2082
139CA2d607
294P2d 474
'56ALC724
30LC¶69842
37LRM2772
1Ap2d607
153S2d1003
'56ALC1079
30LC¶70048
38LRM2262
155S2d 596
'56ALC1228
31LC¶70166
38LRM2541
48W2d525
295P2d 309
'56ALC686
30LC¶69873
37LRM2812
22BdR 98
22BdR131
22BdR145
'56PrAB26
57CR 53
52NwL158
42ABA886
43ABA231
1LE2078n
32A21026s
—198—
(100LE 199)
(76SC 273)
('56ALC
154)
(25LA 693)
(29LC
¶69689)
s349US 943
99LE 1270
75SC 873
s218F2d 948
'55ALC1190
23LA 821
27LC¶68963
s235F2d 209
'56ALC1546
27LA 441
31LC¶70272
s122FS 733
'55ALC 2
23LA 447
26LC¶68779
Arb. & Aw.
Action–
Stay–Stat-
ute–Con-
struction
e233F2d 94
'56ALC901
30LC¶69908
38LRM2019
235F2d 301
'56ALC1399

26LA 736
30LC¶70090
38LRM2378
236F2d 634
241F2d 326
'57ALC403
31LC¶70477
39LRM2384
139FS 18
148FS 523
'57ALC474
31LC¶70483
39LRM2520
Applicable
Law–
Agreement
–Revoca-
tion
d231F2d 763
d236F2d 780
'56ALC1407
31LC¶70179
38LRM2590
d140FS 22
d145FS 17
'56ALC1543
31LC¶70280
38LRM2766
–Remand–
Considera-
tion
233F2d 543
258F2d 470
Arbitra-
tion–Judi-
cial Deter-
minations–
Differ-
ences–
Effect
d356US 535
2LE 961
78SC 899
j356US 550
2LE 969
78SC 907
230F2d 419
241F2d 514
'57ALC384
27LA 820
31LC¶70495
257F2d 315
141FS 225
145FS 18
'56ALC1543
31LC¶70280
38LRM2766
20FRD362
–State
Laws–Ap-
plication
j351US 170
100LE 1064
76SC 735
h230F2d 110
242F2d 538
'57ALC781
32LC¶70590
39LRM2689
139FS 631
'56ALC767
26LA 166

30LC¶69872
37LRM2749
144FS 553
*
— Ca2d —
315P2d 331
40LRM2709
2Ap2d342
155S2d 805
3Ap2d214
159S2d 331
11ArJ 77
11ArJ 127
13ArJ 204
7LLJ428
7LLJ759
8LLJ240
8LLJ784
9LLJ398
9LLJ518
57CR 54
71HLR 8
71HLR131
67YLJ193
67YLJ323
100LE214n
100LE231n
61A2 927n
—247—
(100LE 267)
(76SC 330)
('56ALC
242)
(29LC
¶69711)
(12WHC
750)
s349US 914
99LE 1248
75SC 605
s215F2d 171
'54ALC1425
26LC¶68617
12WHC250
s220F2d 751
s111FS 546
'53ALC554
23LC¶67483
11WHC282
Lab. Rel.
Portal to
Portal–In-
cluded Ac-
tivities
f350US 260
100LE 283
76SC 338
'56ALC241
29LC¶69712
12WHC755
d233F2d 902
'56ALC1018
30LC¶70018
12WHC886
235F2d 624
'56ALC1160
30LC¶70140
13WHC 3
235F2d 646
'55ALC1212
'56ALC1389

28LC¶69435
30LC¶70148
12WHC626
13WHC14
d240F2d 116
'57ALC115
31LC¶70411
13WHC127
j254F2d 12
'58ALC603
34LC¶71452
13WHC627
256F2d 837
'58ALC982
35LC¶71631
13WHC704
*
345Mch 273
76NW 40
'56PrAB28
21A21327s
—260—
(100LE 282)
(76SC 337)
('56ALC
241)
(29LC
¶69712)
(12WHC
755)
s349US 914
99LE 1248
75SC 605
s350US 983
100LE 851
76SC 466
s216F2d 618
'54ALC1801
26LC¶68787
12WHC313
s107FS 369
'53ALC 5
22LC¶67258
11WHC170
Lab. Rel.
Portal to
Portal–In-
cluded Ac-
tivities
d233F2d 902
'56ALC1018
30LC¶70018
12WHC886
235F2d 624
'56ALC1160
30LC¶70140
13WHC 3
235F2d 646
'55ALC1212
'56ALC1389
28LC¶69435
30LC¶70148
12WHC626
13WHC14
240F2d 116
'57ALC115
31LC¶70411
13WHC127
*
345Mch 273
76NW 40
'56PrAB28
21A21327s

—264—
(100LE 285)
(76SC 383)
('56ALC
426)
(29LC
¶69777)
(37LRM
2585)
s108Bd 490
No 81
s350US 819
100LE 733
76SC 49
s219F2d 441
9CtD1023
'55ALC320
27LC¶69007
35LRM2526
Lab. Rel.
Unfair
Practices–
Complaint
–Noncom-
munist Af-
fidavits–
Require-
ment
e352US 149
1LE 204
77SC 156
'56ALC1578
31LC¶70349
39LRM2146
231F2d 769
'56ALC648
29LC¶69817
37LRM2656
e232F2d 518
'56ALC729
30LC¶69912
38LRM2037
235F2d 706
'56ALC1148
30LC¶70147
38LRM2520
f243F2d 307
'57ALC686
32LC¶70621
39LRM2743
141FS 468
'56ALC1201
30LC¶69848
W. & Ph.
Officer
350US 405
100LE 478
76SC 459
d234F2d 777
'56ALC968
30LC¶70041
38LRM2265
*
115Bd 1026
No 164
115Bd 1155
No 185
115Bd 1375
No 217
117Bd 139
No 31

117Bd 321
No 42
i117Bd 325
No 42
117Bd 552
No 81
117Bd 859
No 130
i117Bd 1290
No 168
118Bd 283
No 37
i118Bd 993
No 125
i118Bd 1499
No 209
21BdR 35
21BdR125
21BdR131
9LLJ864
'56PrAB15
67YLJ288
1LE1676n
—270—
(100LE 309)
(76SC 349)
('56ALC
428)
(29LC
¶69779)
(37LRM
2587)
s103Bd 511
No 51
s106Bd 454
No 81
s348US 910
99LE 714
75SC 297
s351US 980
100LE 1495
76SC 1043
s214F2d 462
9CtD 721
'54ALC1337
26LC¶68584
34LRM2484
Lab. Rel.
Act–Inter-
pretation–
Legislative
History–
Consist-
ency
354US 358
1LE 1399
77SC 1143
244F2d 611
147FS 136
–Strike–
Waiting
Period–
Employees
–Status
245F2d 598
'57ALC968
40LRM2213
245F2d 599
'57ALC1001
40LRM2217

246F2d 857
258F2d 147
'58ALC869
35LC¶71616
42LRM2320
Agreement–
Strike–
Right–
Waiver
j352US 302
1LE 344
77SC 341
31LC¶70446
39LRM2296
253F2d 374
'58ALC427
34LC¶71365
Right to
Strike–De-
privation–
Evidence
f352US 284
1LE 335
77SC 331
31LC¶70446
39LRM2296
Statutes
Construc-
tion
j241F2d 214
248F2d 458
154FS 910
W. & Ph.
Any Strike
241F2d 625
'57ALC376
31LC¶70510
39LRM2441
j115Bd 1598
No 260
i117Bd 679
No 105
i117Bd 1121
No 154
117Bd 1131
No 162
i117Bd 1162
No 162
i118Bd 572
No 58
23NJ 105
127A2d 880
'57ALC119
31LC¶70392
39LRM2208
43NJS181
128A2d 25
275Wis 504
82NW 176
32LC¶70626
39LRM2770
21BdR121
22BdR 93
22BdR115
11 ILRR411
8LLJ250
8LLJ364
8LLJ685
8LLJ772
9LLJ 53

9LLJ434
'56PrAB12
67YLJ287
—299—
(100LE 335)
(76SC 400)
('56ALC
423)
(29LC
¶69778)
(37LRM
2581)
s350US 860
100LE 763
76SC 103
s225F2d 417
'55ALC1003
28LC¶69341
36LRM2415
s232F2d 481
'56ALC736
30LC¶69913
38LRM2066
s128FS 128
145FS 39
'56ALC1447
31LC¶70270
38LRM2717
Lab. Rel.
Labor
Manage-
ment Act–
Legislative
History
256F2d 552
'58ALC1135
35LC¶71630
–Welfare
Fund–
Legislative
Purpose
236F2d 116
'56ALC1177
30LC¶70094
38LRM2355
240F2d 265
'57ALC221
31LC¶70445
39LRM2499
162FS 872
'58ALC1108
35LC¶71598
42LRM2200
W. & Ph.
Any Repre-
sentative
of Any Em-
ployees
350US 420
100LE 500
76SC 526
'56ALC508
30LC¶69844
37LRM2731
240F2d 264
'57ALC221
31LC¶70445
39LRM2499
d242F2d 624
'57ALC955
32LC¶70579
39LRM2617
Continued

FIG. 2-14 Portion of a citation index. (From *Shepard's Federal Labor Law Citations.*)

2. Type of immediate user (librarian or customer)

3. Characteristics of the file collection (current and expected size, rate of growth, variety and complexity of subject content, and format of file material)

4. Availability of other existing indexes for the same type of file material

5. Complexity and required accuracy of searches to be conducted (current awareness, comprehensive retrospective searches)

6. Number of searches expected, and their required response time

7. Current user and librarian attitudes toward the existing indexing system and form of display

8. Resources available for developing the system, converting the backlog of material to the new system or new method of display, and maintaining the routine operation

Additional References

American University, *Machine Indexing: Progress and Problems* (American University, Washington, D.C., 1961).

Association of Special Libraries and Information Bureaux, *Proceedings of the International Study Conference on Classification for Information Retrieval,* held at Dorking, England (Pergamon Press, London, May 13–17, 1957).

Atherton, P. A., "File Organization: Principles and Practices for Processing and Maintaining the Collection," Chapt. 2 in *Information Handling: First Principles* (Spartan Books, Washington, D.C., 1962).

Barden, B. R., and B. Denison, *Guide to the SLA Loan Collection of Classification Schemes and Subject Heading Lists,* 5th ed. (Special Libraries Association, New York, 1961).

Bernier, C. L., "New Kinds of Indexes," *Journal of Chemical Documentation,* Vol. 1, No. 1, pp. 62–67 (January 1961).

Bradford, S. C., *Documentation* (Crosby Lockwood & Son, London, 1953).

Coates, E. J., *Subject Catalogues; Headings and Structure* (The Library Association, London, 1960).

Collision, R. L., *Indexes and Indexing,* 2nd ed. (Ernest Benn, Limited, London, 1959).

Dailey, J. E., "Subject Headings and the Theory of Classification," *American Documentation,* Vol. 8, No. 4, pp. 269–274 (October 1957).

Engineering Societies Library, *Bibliography on Filing, Classification, and Indexing Systems for Engineering Offices and Libraries,* ESL Bibliography No. 14 (New York, January 1960).

Fairthorne, R. A., "Basic Postulates and Common Syntax," Chapt. 44 in *Information Retrieval and Machine Translation,* A. Kent, editor (Interscience Publishers, New York, 1960).

Herdman, M. M., *Classification: An Introductory Manual,* 2nd ed. (American Library Association, Chicago, 1947).

Kelley, G. O., *The Classification of Books: An Inquiry Into Its Usefulness to the Reader* (H. W. Wilson Co., New York, 1937).

Kent, A., editor, *Information Retrieval and Machine Translation,* Parts 1 and 2 (Interscience Publishers, New York, 1960).

Lloyd, G. A., "A 'New Deal' for Universal Classification," *Revue de la Documentation,* Vol. 27, No. 4, pp. 149–153 (November 1960).

MacDonald, M. R., "Book Catalogs and Card Catalogs," *Library Resources and Technical Services,* Vol. 6, No. 3, pp. 217–222 (Summer 1962).

Markus, J., "State of the Art of Published Indexes," *American Documentation,* Vol. 13, No. 1, pp. 15–30 (January 1962).

Palmer, B. I., and A. J. Wells, *The Fundamentals of Library Classification* (Allen & Unwin, London, 1951).

Perry, J. W., and A. Kent, *Documentation and Information Retrieval: An Introduction to Basic Principles and Cost Analysis* (The Press of Western Reserve University and Interscience Publishers, New York, 1957).

Richmond, P. A., "Hierarchical Definition," *American Documentation,* Vol. 11, No. 2, pp. 91–96 (April 1960).

Richmond, P. A., "Recommendations for Further Study in Classification," *American Documentation,* Vol. 12, No. 1, pp. 35–37 (January 1961).

Richmond, P. A., "Some Aspects of Basic Research in Classification," *Library Resources and Technical Services,* Vol. 4, No. 2, pp. 139–149 (Spring 1960).

Salton, G., "Manipulation of Trees in Information Retrieval," *Communications of the Association for Computing Machinery,* Vol. 5, No. 2, pp. 103–114 (February 1962).

Samore, T., "Form Division in L.C. and D.C. Classification Schemes," *Library Resources and Technical Services,* Vol. 6, No. 3, pp. 243–246 (Summer 1962).

Shera, J. H., and M. E. Egan, *The Classified Catalog: Basic Principles and Practices* (American Library Association, Chicago, 1956).

Stevens, N. D., *A Comparative Study of Three Systems of Information Retrieval* (Rutgers University Press, New Brunswick, New Jersey, 1962).

Taube, M., "Specificity in Subject Headings and Coordinate Indexing," *Library Trends,* October 1952, pp. 219–223.

Taube, M., and H. Wooster, *Information Storage and Retrieval* (Columbia University Press, New York, 1958).

Tauber, M. F., editor, *The Subject Analysis of Library Materials* (School of Library Science, Columbia University, New York, 1953).

Vickery, B. C., *Classification and Indexing in Science,* 2nd ed. (Academic Press, New York, 1959).

Vickery, B. C., *On Retrieval System Theory* (Butterworth, Washington, D.C., 1961).

Weeks, B. M., *How to File and Index,* rev. ed. (Ronald Press, New York, 1951).

Whaley, F. R., "The Manipulation of Non-Conventional Indexing Systems," *American Documentation,* Vol. 12, No. 2, pp. 101–107 (April 1961).

The Statistical Nature of English Words
Methods for Systematically Abbreviating English Words and Names
Prime Number Coding
Superimposed Coding

3

Coding: The indexing shorthand

In addition to establishing the over-all organization of the file collection, the system designer must decide how the classification or indexing information is to be represented, whether on paper or in some machine-readable form. In most cases, coding is a shorthand representation of the indexing information, and can help system operation by reducing clerical effort and storage requirements, and by streamlining the search and retrieval process. The choice of coding method becomes critical if automatic or semi-automatic processing or handling of the file or indexing information is to be used, and this special topic of machine language representation is examined in Chapt. 4. There are, however, several special and useful coding topics that have been described previously only in widely scattered sources. These less-well-known coding techniques and background information have been collected into this chapter. Other coding topics are discussed in Chapts. 5, 6, and 7 as they relate to specific types of equipment or media. Further information is available from the sources listed at the end of the chapter.

THE STATISTICAL NATURE OF ENGLISH WORDS

Frequency distributions for the alphabetic letters in English text material have been determined by at least six investigators.[1] Two studies in particular concentrated on the statistical nature of people's names and technical subject words.[2] A summary of many of the results of these studies is presented here.

Individual Letter Usage

Table 3-1 gives five different rankings of frequency of occurrence of initial letters. Three different studies of continuous text material show relatively close correlation and agree that T, A, and O are the three most frequent characters for initial letters. Two different studies of subject words also show relatively close correlation and agree that S, P, and C are the three most frequent characters for

[1] Pratt, F., *Secret and Urgent, The Story of Codes and Ciphers* (Blue Ribbon Books, Garden City, New York, 1942).

Ohaver, M. E., *Cryptogram Solving* (Stoneman Press, Columbus, Ohio, 1933).

Cox, G. J., et al., "Recent Developments in Keysort Cards," *Journal of Chemical Education*, Vol. 24, pp. 65–70 (February 1947).

Smith, L. D., *Cryptography—The Science of Secret Writing* (W. W. Norton Co., New York, 1943).

Gaines, H. F., *Cryptanalysis* (Dover Publications, New York, 1956).

Griffith, R. T., "The Minimotion Typewriter Keyboard," *Franklin Institute Journal*, Vol. 248, No. 5, pp. 399-436 (November 1949).

[2] Bourne, C. P., and D. F. Ford, "A Study of the Statistics of Letters in English Words: *Information and Control*, Vol. 4, No. 1, pp. 48–67 (March 1961).

Ohlman, H., "Subject Word Letter Frequencies with Applications to Superimposed Coding," *Proceedings of the International Conference on Scientific Information*, Vol. 2, pp. 903–916 (National Academy of Sciences, Washington, D.C., 1959).

TABLE 3-1

Comparative Ranking by Frequency of Ocwrrence of Initial Letters in English Words

Subject Words (Bourne-Ford)	Subject Words (Ohlman, p. 92)	Continuous Text (Pratt, p. 258)	Continuous Text (Ohaver, p. 27)	Continuous Text (Smith, p. 153)
S	C	T	T	T
P	S	A	A	O
C	P	O	O	A
T	A	S	S	W
A	B	W	H	B
D	M	I	I	C
M	T	H	W	D
R	R	C	C	S
B	E	B	B	F
I	F	F	P	M
E	D	P	F	R
F	G	M	D	H
H	H	R	M	I
L	I	E	R	Y
G	L	L	E	E
W	N	N	Y	G
N	W	D	N	L
O	O	U	L	N
V	V	G	U	P
J	U	Y	G	U
U	J	J	V	J
K	K	V	K	K
Q	Q	Q	J	
Y	Y	K	Q	
Z	Z	X	Z	
X	X	Z	X	

initial letters. Taken as a group, the rankings for subject words are different from those for continuous text. This difference would be expected, primarily because of the large number of prepositions and other small words (e.g., a, and, in, is, the, to, with, etc.) that appear in continuous text.

Table 3-2 shows that four different studies of initial letters in people's names were in relatively close agreement. All the studies showed that B, S, M, H, and C are the most frequent characters for initial letters.

The three studies of continuous text agreed that E, S, D, T, and N were the most frequent characters for terminal letters, and were in relatively close agreement for the remainder of the characters. The rankings of terminal letters in the study of subject words were distinctly, but not radically, different from those for continuous text. All the studies agreed that E was the most common terminal letter.

Table 3-3 lists data from six different studies of the composite rankings of letter usage. Composite rankings are based on the relative frequency of occurrence of letters in collections of words or in text and do not consider the position of the letters within the words. The four studies of continuous text material showed remarkably close agreement, and all showed that E, T, A, O, and N were the most frequent letters. The rankings for subject words and for proper names differed from each other and from the rankings for continuous text material.

TABLE 3-2

Comparative Ranking by Frequency of Occurrence of Initial Letters in Proper Names

Bourne-Ford [a]	Bourne-Ford [b]	Ohlman, p. 93	Cox et *al.*, p. 69
B	S	S	S
S	B	B	B
M	M	M	M
H	H	H	H
C	C	C	C
W	W	D	W
R	R	G	R
L	G	K	P
P	L	L	L
G	P	R	G
D	D	P	D
K	F	W	F
F	K	A	K
T	T	F	T
A	A	T	A
J	J	E	E
E	E	N	J
N	N	V	N
O	O	J	V
V	V	O	O
Y	I	I	Y
I	Y	Z	I
Z	Z	U	Z
U	U	Y	U
Q	Q	Q	Q
X	X	X	X

[a] Based on a detailed study of 8207 names.

[b] Based on an independent, but less-detailed study of approximately 63,000 names.

TABLE 3-3

Comparative Ranking by Composite [a] Frequency of Occurrence of Letters in English Words and Proper Names

Subject Words (Bourne Ford)	Continuous Text (Pratt, p. 252)	Continuow Text (Ohaver, p. 25)	Continuow Text (Griffith, p. 426)	Continuous Text (smith, p. 153)	Proper Names (Bourne Ford)
E	E	E	E	E	E
I	T	T	T	T	A
R	A	A	A	O	R
O	O	O	O	A	N
A	N	N	N	N	L
T	R	I	I	I	O
N	I	S	S	R	I
S	S	R	R	S	S
L	H	H	H	H	T
C	D	L	L	D	H
P	L	D	D	L	D
M	F	C	C	C	M
D	C	U	U	W	C
U	M	P	M	U	B
H	U	F	F	M	G
G	G	M	Y	F	U
Y	Y	W	W	Y	W
B	P	Y	G	G	Y
F	W	B	P	P	J
V	B	G	B	B	K
K	V	V	V	V	P
W	K	K	K	K	F
X	X	Q	X	X	V
Z	J	X	J	Q	Z
J	Q	J	Q	J	X
Q	Z	Z	Z	Z	Q

[a] These are all composite samples which did not consider the position of the letters within the words.

The only point of complete agreement in all six studies was that E was the most common letter. A more graphic comparison of the composite distributions for subject words and proper names is given in Fig. 3-1, which shows the actual frequencies of usage as well as the rankings.

Figure 3-2 illustrates the rankings and frequency distributions for the characters of subject words in a particular letter position (i.e., the frequency distribution for the initial letter of the word, the second letter of the word, etc.). Some interesting patterns are shown in this figure. For example, the distributions of the first, second, and possibly the third, letter positions have relatively unique patterns, and those of the remaining letter positions have generally the same pattern. In the second letter position, the vowels are the most common characters. The ranking is E, O, A, I, R, U. Excluding the R, the group of five vowels accounts for over 61 percent of the letter usage in that position.

Figure 3-3 illustrates the rankings and frequency distributions for the characters of proper names within a particular letter position. As in the case of the subject words, the distributions of the first and second letter positions have relatively unique

TABLE 3-4

Comparative Ranking by Composite [a] Frequency of Occurrence of the Most Common [b] Bigrams in English Words and Proper Names

Continuous Text (Pratt, p. 260)	Military Telegraph (Pratt, p. 260)	Continuous Text (Gaines, p. 218)	Continuous Text (Smith, p. 153)	Continuous Text (Ohaver, p. 26)	Subject Words (Bourne Ford)	Proper Names (Bourne-Ford)
TH	TH	TH	TH	TH	ER	ER
HE	AN	HE	ER	HE	ON	AR
AN	RE	AN	ON	ER	TI	AN
RE	HE	IN	AN	IN	IN	ON
ER	ON	ER	RE	AN	TE	LE
IN	IN	RE	HE	ON	AT	EN
ON	ER	ES	IN	RE	RA	LL
AT	AT	ON	ED	AT	RO	HA
ND	TO	EA	ND	ED	IO	EL
ST	OU	TI	HA	ST	AN	NE
ES	OR	AT	AT	ND	OR	RO
EN	ND	ST	EN	ES	IC	AL
OF	TA	EN	ES	HA	RE	IN
TE	EN	ND	OF	OF	TR	MA
ED	ES	OR	OR	EN	AL	RI
OR	NT	TO	NT	OR	EN	CH
TI	TI	NT	EA	AS	AR	LI
HI	HA	ED	TI	NT	NT	RE
AS	ST	IS	TO	TI	LE	TH
TO	RO	AR	IT	TO	ET	BE
AR	AR	OU	ST	IT	NE	AM
OU	CO	TE	IO	EA	RI	ES
IS	TE	OF	LE	NG	ST	IL
IT	IT	IT	IS	OU	IT	RT
LE	ME	HA	ON	AR	ES	RD
NT	TT	SE	AR	ET	CO	RA
RI	OM	ET	AS	HI	LA	LA
SE	VE	AL	DE	TE	DE	OR
HA	UR	RI	RT	RO	ME	IC
AL	ED	NG	VE	NE	LI	TE

[a] These rankings did not consider the position of the letter pairs within a word.

[b] The bigrams which consisted of a letter and a space were not considered for these rankings.

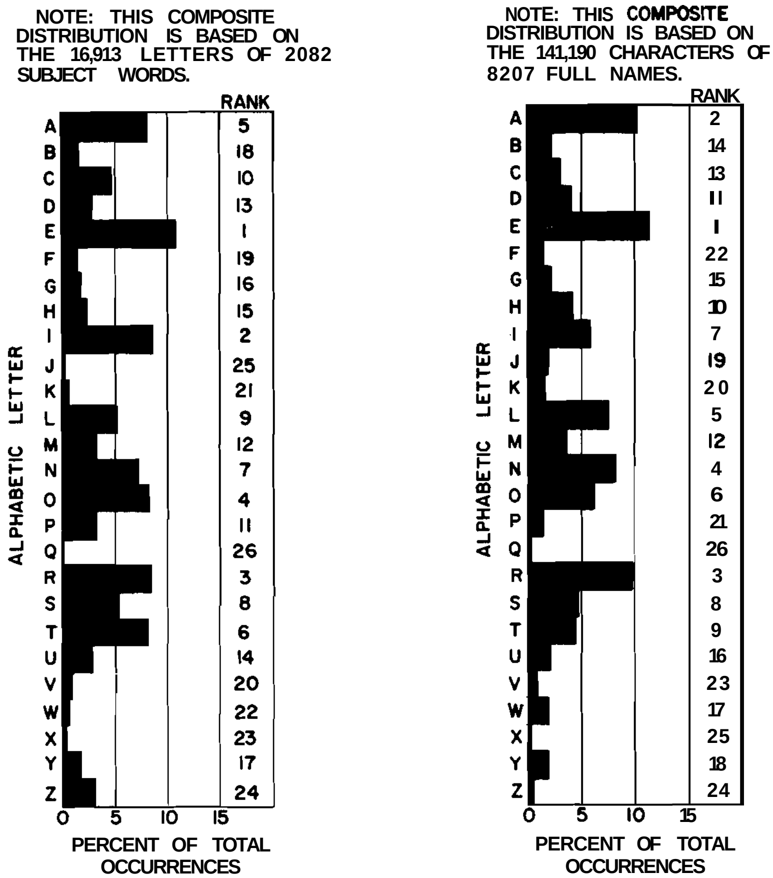


FIG. 3-1 Composite distributions and rankings for the alphabetic letters in subject words and people's names.

patterns, while those for the remainder of the letter positions have much the same pattern. The vowels dominate the second letter position of proper names, as they did with subject words. In this case, the vowels (ranked A, O, E, I, U), account for over 70 percent of the character usage in the second letter position. There is general agreement between the patterns for corresponding letter positions of subject words and proper names.

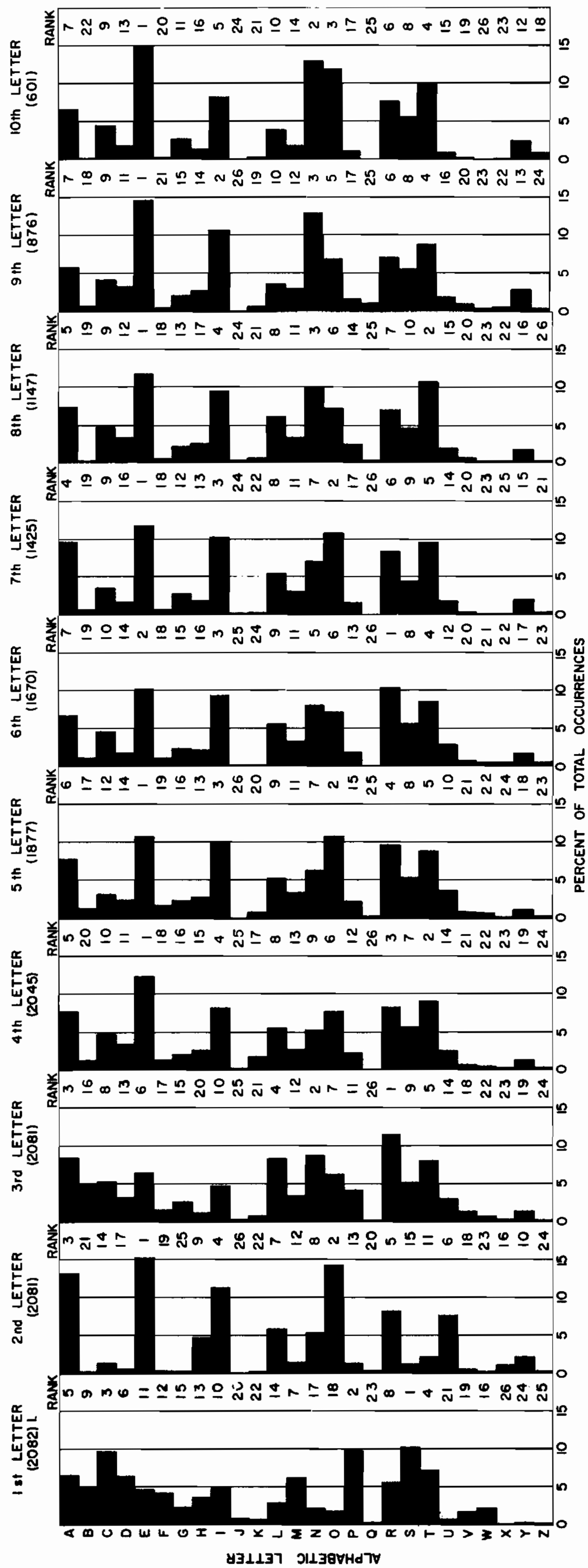
Bigram (letter Pair) Usage

Table 3-4 lists data from seven different studies of rankings of bigram occurrences. The five different studies of continuous text and telegraph text agreed in many respects (e.g., that TH was the most common bigram). The bigram rankings for subject words and for proper names were different (except in agreeing that ER was the most common bigram) and somewhat different from any of the other five rankings. However, because of the large number of bigram possibilities it is more difficult to make a quantitative judgment of the degree of similarity between bigram rankings than it is for the ranking of single alphabetic characters. The frequency of occurrence of various bigrams in subject words and proper names and the composite bigram rankings for subject words and proper names are reported in the literature.[3]

Distribution of *Subject* Word lengths and Proper *Name Lengths*

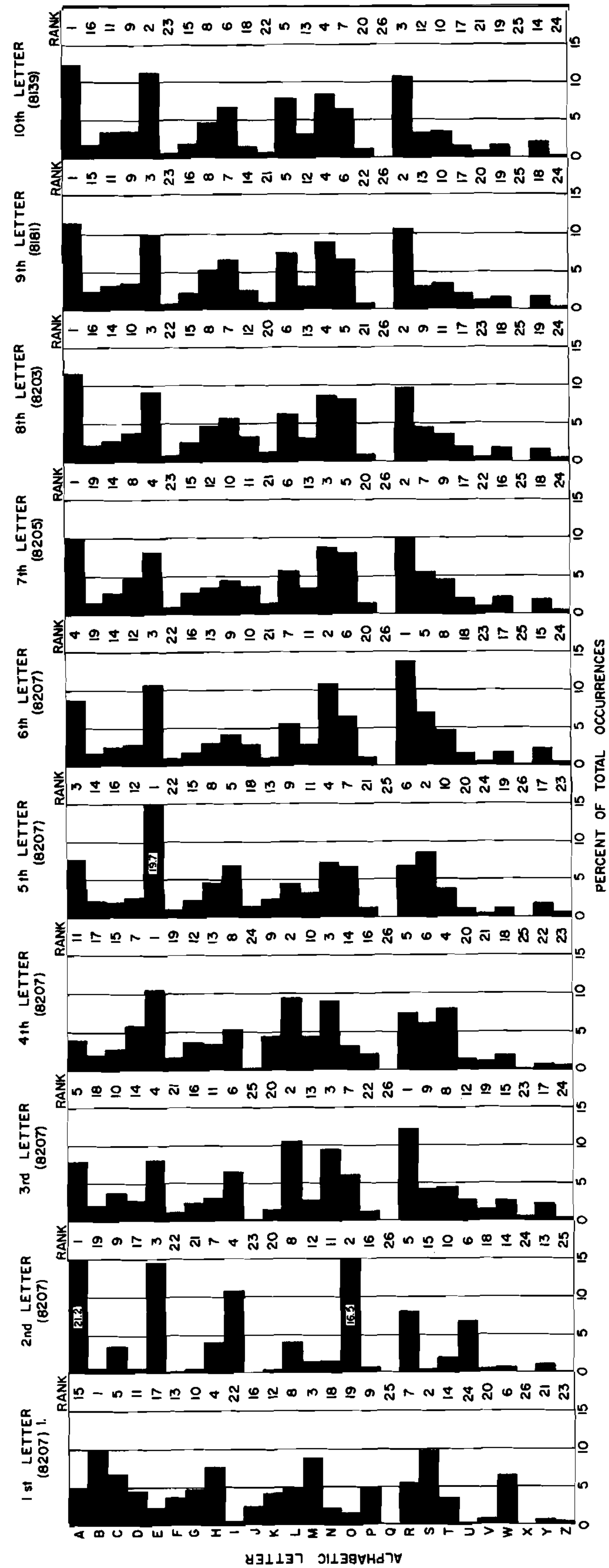
The distributions of subject word lengths and of proper name word lengths are illustrated in Figs. 3-4 and 3-5, respectively. The subject words rep-

[3] Bourne and Ford, *Information and Control,* article *cited earlier in this chapter!* fn. *2.*



1. THE NUMBER IN PARENTHESIS REFERS TO THE NUMBER OF NAMES IN THE SAMPLE.

FIG. 3-2 Frequency of occurrence and rankings for the letters of subject words by letter position.



1. THE NUMBER IN PARENTHESIS REFERS TO THE NUMBER OF NAMES IN THE SAMPLE.

FIG. 3-3 Frequency of occurrence and rankings for the letters of proper names by letter position.

resented single-word descriptors (e.g., "magnetic," "optical") which were used to index a collection of technical documents in a coordinate indexing system. The names consisted of the last name as well as some combination of the first and middle (or more) names, and their prefixes (in that order).

Statistical Summary

In general, the statistical characteristics of usage of the alphabetic letters are markedly different for three different categories of English words—continuous text material, proper names, and subject words. Several independent studies of continuous text material show close agreement for a number of characteristics, suggesting that the reported data are representative of the general nature of continuous English text material.

In subject words (two studies), S, P, and C were the three most frequent initial letters. In continuous text (three studies), T, A, and O were the three most frequent initial letters. In proper names (four studies), B, S, M, H, and C were the most frequent initial letters.

In subject words (one study), E, N, R, and S were the most frequent terminal letters. In continuous text (three studies), E, S, T, and D were the most frequent terminal letters.

In subject words (one study), E, I, R, and O were the most frequent letters. In continuous text (four studies), E, T, A, O, and N were the most frequent letters. In proper names (one study), E, A, R, and N were the most frequent letters.

In subject words and proper names (one study each), the first, second, and possibly the third letter positions had relatively unique letter distribution patterns, and the remaining letter positions had patterns that closely resembled each other.

In subject words (one study), the most frequent bigrams were ER, ON, TI, and IN. In continuous text and telegraph text (five studies), TH, AN, HE, and RE were some of the most frequent bigrams. In proper names (one study), the most frequent bigrams were ER, AR, AN, and ON.

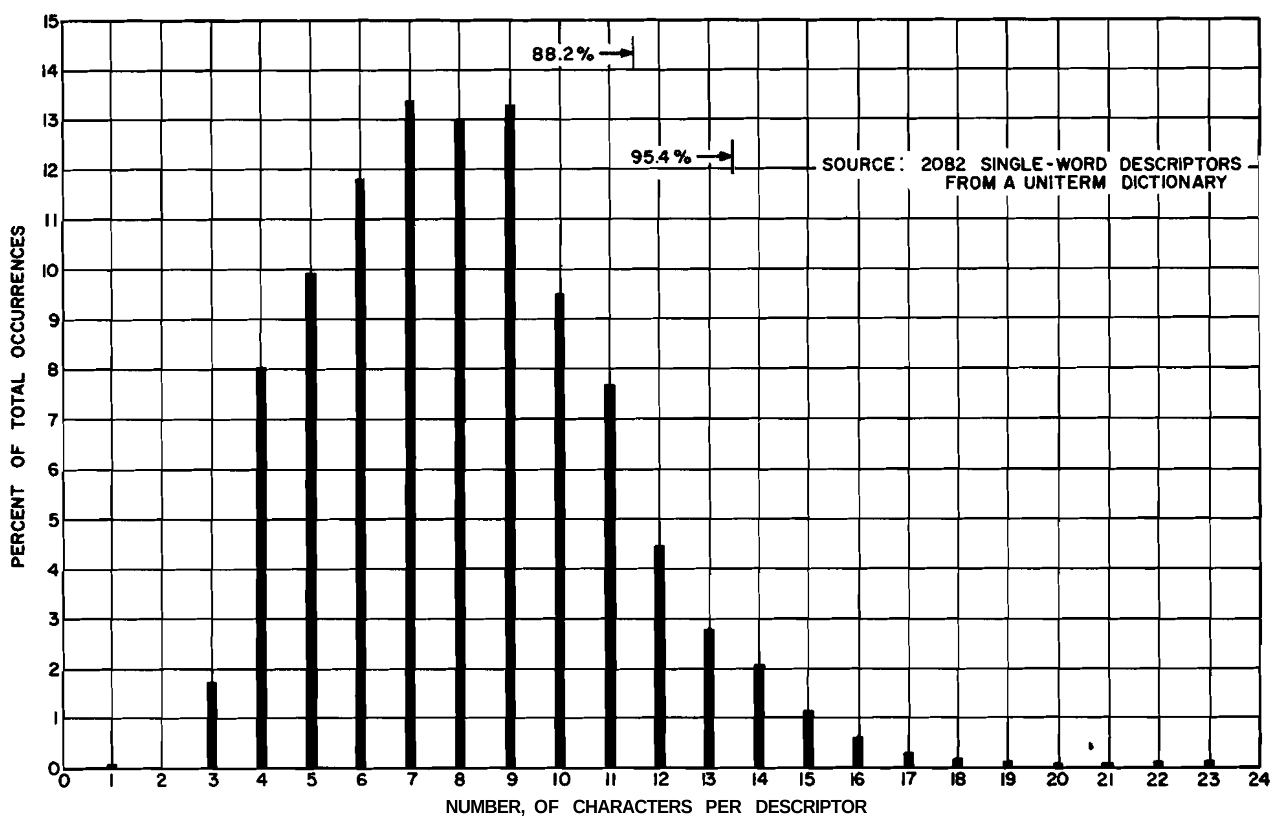


FIG. 3·4 Distribution of subject word lengths.

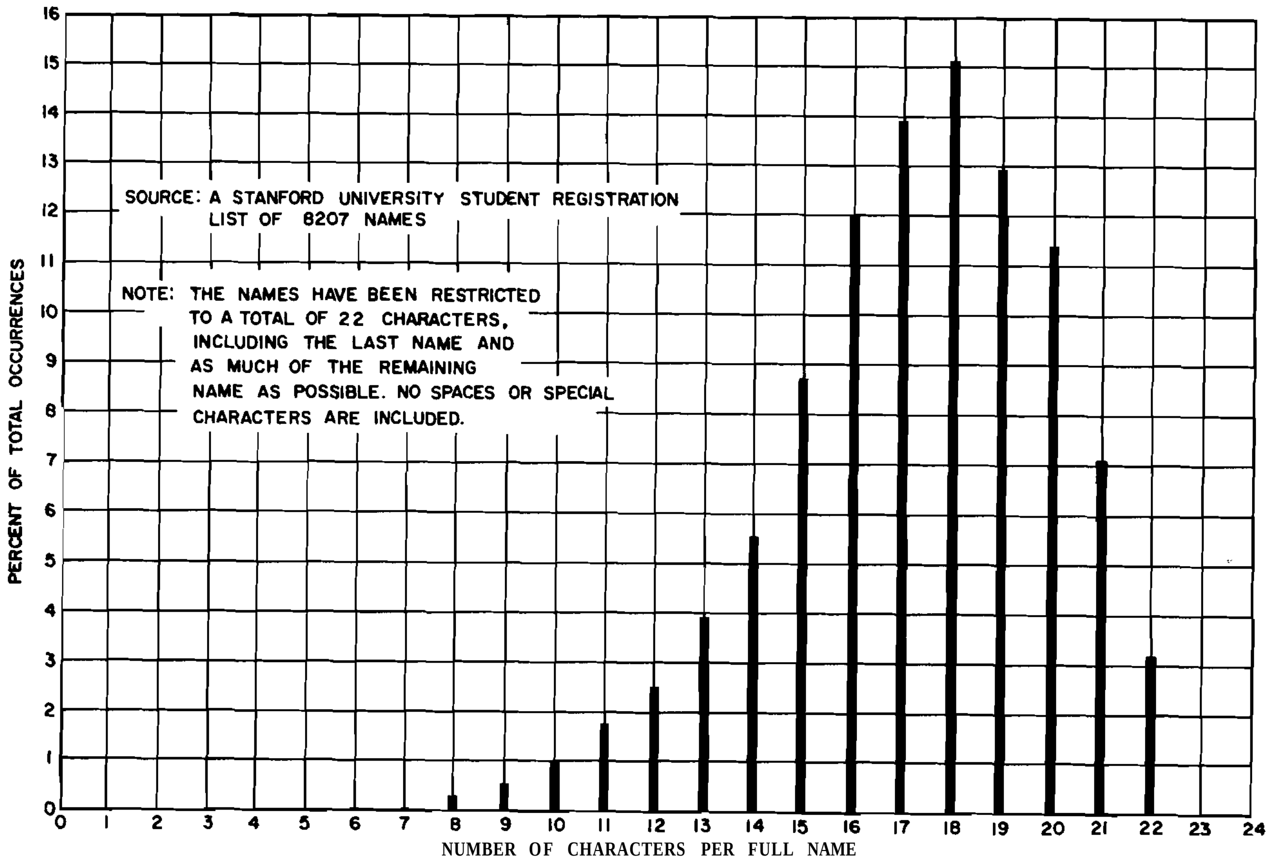


FIG. 3-5 Distribution of proper name lengths.

In single-word descriptors, one study showed that over 95 percent of the vocabulary could be represented by 13 characters. For proper names, one study showed that approximately 95 percent of the names could be represented by 25 characters.

METHODS FOR SYSTEMATICALLY ABBREVIATING ENGLISH WORDS AND NAMES

There are many instances in which it is useful to abbreviate [4] English words such as people's names, places, street addresses, proper nouns, or continuous text material. Common subjects for abbreviation are addresses in telephone directories, customer account number identifications in some data processing systems, and commercial teletype or ham radio vocabularies. There will always be a need for abbreviation or coding schemes to make efficient use of the storage media in computers, punched card systems, and other storage and processing systems. Abbreviation techniques may also improve the efficiency of communication systems by decreasing the number of characters transmitted. If the abbreviations are to be made automatically or semiautomatically, they must be obtained by some systematic method.

Many methods have been suggested for abbreviating words, and several of them are described in the following sections. A few of these schemes have been empirically tested with representative data to obtain meaningful performance comparisons.

Some abbreviation studies have been concerned primarily with the subject words or descriptors used with documentation systems. Other studies have been concerned with the abbreviation of names or

[4] This discussion refers to the "abbreviation" of words, but actually it is concerned with "coding" the words, or transforming the source words into new patterns. It happens that in satisfying some of the objectives of code design the word is abbreviated. However, some of the schemes and transformations described in the following sections bear little resemblance to what might normally be considered as abbreviation.

the derivation of customer account numbers, and a few studies have been concerned with continuous text material. The objective of most of these studies has been to develop methods for the systematic coding of words which satisfy the following requirements:

1. Each word should be coded to require as little storage space as possible.
2. Each word should retain the same degree of discrimination and uniqueness that it had in the original sample. (For example, if there were 2082 unique words in the original list of words or names, hopefully there will still be 2082 unique items after abbreviation.)
3. If possible, each word should retain some mnemonic similarity to the original word.
4. It would be a useful feature if the abbreviated word could be systematically transformed back into the original word when desired.
5. It would be a useful feature if the abbreviated words could be used as a basis for sorting the original words in nearly correct alphabetic order.

These are idealized requirements which have not been met in full by any of the methods described in the following sections.

The abbreviation procedure may or may not rely on prior detailed knowledge of the body of words to be abbreviated. A method based on prior detailed knowledge may be called an assigned code system, and a method not requiring such knowledge is called a derived code system. From the standpoint of uniqueness and storage space, one of the most efficient schemes would be an assigned code method that used a previously established table of code numbers or abbreviations for each word that would subsequently be encountered. The scheme would require the initial development of a table and continued reference to it. Derived code systems would generally accept and abbreviate any word presented to the system, without the use of detailed tables generated in advance.

As illustrated in Table 3-5, the 26 characters of the English alphabet offer a large number of coding possibilities, so that a few alphabetic characters are capable of providing a large degree of discrimination. Even if the non-pronounceable letter combinations were removed, there would still be a large number of code combinations. The use of a few alphabetic characters in an automobile license number, for example, provides a large number of possible codes with just a few letters. There has even been a compilation and publication of 1805 three-letter words and 19,005 four-letter words which have the special property of being self-demarcating.[5] That is, the words are serially unique, disjunctive in combination, and reasonably pronounceable. A computer program has even been written to generate pronounceable three-letter and four-letter words.[6]

TABLE 3-5

Number of Unique Codes Possible with a Series of Alphabetic Letters

Number of Characters Used in the Abbreviation	Number of Unique Codes Possible
1	26
2	676
3	17,576
4	456,976
5	11,821,376
n	26^n

Derived Coding Methods

The derived coding methods generate a code directly from the source material without the necessity for table look-ups.[7] In nearly all the methods described in this section, the initial letter is retained, and the scheme is not executed if the word length is initially less than or equal to the desired length of the code word. Each of 13 different methods is briefly described below.

Simple Truncation of the Right End. Starting from the right end of the word, drop off letters until the

[5] Luhn, H. P., Self-Demarcating Code Words, 2nd ed. (IBM Corp., New York, 1953).

[6] Leiner, A. L., and W. W. Youden, "A System for Generating Pronounceable Names Using a Computer," *Journal of the Association for Computing Machinery*, Vol. 8, No. 1, pp. 97–103 (January 1961).

[7] In this text, a "table look-up" is defined as the operation of looking up some item in a table or index, given a particular search key. For example, to look for a specific telephone number in a telephone directory, given a person's name, would be considered a table look-up operation.

required word size is obtained. No exceptions or other operations are allowed. This very simple operation is often used to derive some library call numbers.

Simple Truncation of the Left End. This process is the same as the simple truncation of the right end, except that the letters are dropped from the left end of the word.

Elimination of Vowels. Starting from the right end of the word eliminate vowels (a, e, i, o, u) until the desired word length is reached. If the word cannot be sufficiently shortened by this method, use simple truncation of the right end to reach the desired length after the vowels are eliminated. The common method of manual note taking referred to as "speedwriting" uses this type of coding (plus the elimination of other characters) in order to reduce the amount of material to be written.

Selective Drop-Out *by* Letter Position. Starting from the left end of the word, eliminate every nth letter. (Each different n constitutes a different technique and results in different efficiencies.) The use of this technique to reduce a word to three characters by dropping out every second character (i.e., n = 2) is shown in the two examples below.

First pass	A~~B~~L~~A~~T~~I~~O~~N~~	A~~C~~C~~E~~P~~T~~A~~N~~C~~E~~
Second pass	A~~L~~TO	A~~C~~P~~A~~C
Final result	ATO	APC

Selective Drop-Out *by* a Single Ranking of Letter Usage. Using an empirical ranking of the composite[8] frequency of usage of the 26 alphabetic characters, such as the ones given in Table 3-3, eliminate the most common letters until the desired word length is reached.[9] In case two letters are equally common, remove the right-hand letter first. Because of the different source material used in previous empirical studies, the published frequency distributions may not accurately represent the statistical nature of the new data to be abbreviated or processed. It may be necessary to test or determine the frequency distribution of the new input data before using any of the abbreviation schemes that require a frequency distribution.

Selective Drop-Out *by* Separate *Rankings* of Character Usage for Each Letter Position. Using a separate empirical ranking of letter usage for each letter position, examine each word to determine the respective ranking of each of the characters present. Then remove the letters in the order of their frequency until the desired word length is obtained. In case two letters have equal rank, remove the right-hand letter first.

It has been found that there is a marked similarity in the ranking and distribution of most of the characters, regardless of their letter position (see Figs. 3-2 and 3-3). However, for both subject words and proper names the distributions of the first two, and possibly three, letter positions are each markedly different from the distributions of all the other letter positions. In particular, the frequency of vowels in the second letter position suggests that this position does not provide a great deal of discrimination.[10]

Selective Drop-Out *by* a Single Ranking of *Bigram* (Letter Pair) Usage. Using an empirical ranking of the composite frequency of occurrence of bigrams (i.e., ignoring the position of bigram occurrence within the words), examine each word to determine the ranking of all the possible letter pairs (adjacent letters only). The elimination process considers the association of each letter with the neighboring letters and rejects the letters which make up the most common bigrams. With the exception of the initial letter (which is retained by policy) each letter contributes to two different bigrams. The total ranking of the frequencies of these two bigrams is the selection criterion for that particular letter. Each letter is evaluated in this way to derive its index or measure of discriminative power. Those with the least discriminative power are removed to

[8] That is, an average frequency of usage, regardless of the letter positions. There are distinct differences in the character frequency distributions for the various letter positions (Figs. 3-2 and 3-3).

[9] Luhn, H. P., "Superimposed Coding with the Aid of Randomizing Squares for Use in Mechanical Information Searching Systems," Chapt. 23 in *Punched Cards—Their Application to Science and Industry*, 2nd ed., Casey et al., editors (Reinhold Publishing Corp., New York, 1958).

[10] This has been recognized by other workers, notably Cox, Casey, and Bailey, who have designed punched card codes (for coding proper names) that either ignored the second letter, or provided separate coding spaces for each vowel and one or two additional code spaces to take care of the remaining 21 characters (Cox, G. J., et al., cited previously, fn. 1).

arrive at the desired word length. The process is shown by the following example, reducing the word ACETYLENE to four characters:

Bigram rankings	66	58	25	108	164	24	20	26	2	
Letter	A	C	E	T	Y	L	E	N	E	(space)
Letter index		124	83	133	272	188	44	46	28	
Final result	A	T	Y	L						

It would also be possible to operate with **trigrams** or larger groups of letters, and consideration could be given to letter pairs that are separated by one or more letters. It would also be interesting to examine the use of a separate **bigram** ranking for each letter position.

Truncation after Elimination of the Second Character. Eliminate the second character of the word, counting from the left side of the word. Then perform simple truncation, starting at the right end of the word. No exceptions or other operations are allowed. This modification of the simple truncation scheme should benefit by omission of the second letter with its weak discriminative power.

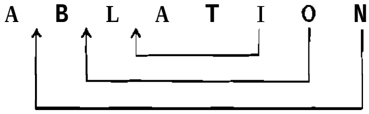
The Use of a Check Digit (or Letter). The discriminative power of any of the preceding coding schemes can probably be enhanced by using a check digit or check letter to describe the letters that have been **eliminated.**[11] For example, an abbreviation technique may reduce COMPUTERS and COMPUTATION to the same root, COMPUT, with a resulting loss in discrimination. Using a check letter, the original words would abbreviate to COMPU *, the letter in the asterisk position generally being different for each original word. There are several ways to generate a check letter: (1) add the numbers that represent the alphabetic position of the eliminated letters (**e.g.,** A = 1, B = 2, . . . , Z = 26); (2) add the numbers that represent the frequency **rankings** of the eliminated letters (**e.g.,** E = 1, A = 2, . . . , Q = 26 for names) ; and **(3)** count the number of eliminated letters. If the number resulting from any of these operations is larger than 26, then cast out 26 as many times as possible; the remainder (less than 26) is then used as the entry in a table of check letters. The table of check letters could be a simple listing of the alphabet. One method for deriving a check digit is shown in the following example, in which the word ABLATION is reduced to a total of four characters by simple truncation and the addition of a check letter:

	1+20+9+15+14=59=26+26+7=H							
Original word	A	B	L	A	T	I	O	N
Final abbreviation	A	B	L	H				

The table for the eliminated letters uses A = **1**, . . . , Z = 26; but the table for the derivation of the check letters uses values A = 0, . . . , **Z = 25** since it is possible to obtain the value zero if the first number exactly equals a multiple of 26.

Shuffle and Truncation. Shuffle the letters of the word, then drop the letters from the right end until the desired word length is obtained. The **shuffle** consists of a simple folding of the letters within the word, as in the following example, in which the word ABLATION is reduced to four letters. Different types of folding patterns can be **used.**

A B L A T I O N = ANBO~~LIAT~~ = ANBO



Elimination *of* Miscellaneous Redundancies. Some benefit may result from operations like the deletion of U after Q, or of one of a pair of the same letters (**e.g.,** tt, mm, nn), or of the second vowel of a double vowel (**e.g.,** ae, io, oe). These possibilities for reduction are not frequent enough to be the basis of a complete technique, but taking advantage of them whenever they do occur can enhance other techniques.

Arithmetic Manipulations. In some types of computer equipment it is possible to do such things as squaring the original word and using some of the digits of the result as the abbreviated word. There would be no resemblance to the original word, but the transformation might yield a relatively unique expression. The abbreviations would probably be numbers, or mixed alphanumeric and special characters. Other types of arithmetic operations could also be used.

Soundex Code. The conventional **Soundex** system for manual files converts a name to a code word having one alphabetic character and three numeric **characters.**[12] It is possible to generate larger code

[11] **Check digits are generated and used in many ways with data processing systems. However, the application of check digits to word abbreviation to represent the omitted letters is relatively new. (See H. P. Luhn, Chapt. 23 in Casey et al., mentioned earlier in this chapter, fn. 9.)**

[12] **Remington Rand,** ***Soundex—Foolproof Filing System for Finding Any Name in the File,*** **Brochure LBV809 (undated).**

Remington Rand, ***Idem Sonans Says It's Legal (Soundex),*** **Brochure LBV528 (undated).**

words, but this is seldom done. The conversion rules are these:

1. Always retain the initial letter of each name.
2. Drop out A, E, I, O, U, Y, W, and H.
3. Assign the following numbers to the remaining similar-sounding sets of letters:

B, F, P, V = 1
C, G, J, K, Q, S, X, Z = 2
D, T = 3
L = 4
M, N = 5
R = 6
Insufficient consonants = 0
(e.g., DARLINGTØN = D645)

4. Special cases:
i. If there are insufficient letters, fill out with zeros (e.g., MØRAN = M650).
ii. Drop out the second letter in a letter pair (e.g., KELLEY = K400).
iii. Drop out adjacent equivalent letters (e.g., JACKSØN = 5250).
iv. Drop out adjacent equivalents of the first letter (e.g., LLØYD = L300).

Most of the methods of shorthand note taking utilize such phonetic coding (without numbers) to reduce the bulk of material to be written. These shorthand methods obtain further condensation by using a family of special graphic symbols to represent the phonetics.

Comparison of Derived Coding Methods

In most coding situations, the major objective is to condense the words as much as possible while maintaining a maximum amount of discrimination between the words. Success in meeting this objective can be measured more accurately than for any of the other objectives mentioned previously. Tests have been made on several of these techniques by abbreviating the test material to a given number of characters, then counting the number of unique entries that remain after the abbreviation process.[13] In this way the degree of uniqueness was measured for various amounts of abbreviation. These data provide a measure of the performance of each scheme, and can be plotted to form summary curves of the operating characteristics of each abbreviation scheme. In practice, the codes should also be evaluated on such criteria as ease of use and derivation.

The data base used for the experimental comparisons shown is Figs. 3-6 and 3-7 consisted of a list of 2082 different single-word Uniterms (e.g., "magnetic," "optical") that were used in an operational coordinate indexing system to index a collection of technical documents. The full and correct spelling of the words was used for all the tests.

The data base used for the comparisons in Fig. 3-8 was a list of the 8207 names of the entire 1959 student registration of Stanford University. In its original application for student registration, each of the names was restricted to a field of 25 characters and consisted of at least the last name and some combination of the first and middle (or more) names and their prefixes. The full names were condensed by removing all spaces and special characters (e.g., hyphens) and compressing them to form single sequences of 22 or fewer characters.

Figures 3-6 and 3-7 show the results of tests on the sample of subject words; Fig. 3-8 shows the results of tests on the sample of proper names. These three figures show, for a given number of code characters used, the ratio (expressed as a percent) of the number of unique entries after abbreviation to the number of unique entries before abbreviation. In general, none of the tested techniques produced abbreviations that could be transformed back into the original words, and none of the techniques produced abbreviations that bore a strong resemblance to the original word. With the exception of the statistical data describing the nature of the words to be abbreviated, no table look-up or prior knowledge of the input data was required.

The familiar techniques of simple truncation and vowel elimination produced the poorest results. Schemes that took advantage of the statistical nature of the words (i.e., drop-out by a composite frequency distribution, drop-out by a separate frequency distribution for each letter position, and drop-out by bigram rankings) produced, better results. The shuffle scheme produced mediocre results, and the technique of selectively dropping the nth letter produced some good results and some bad results. In the two basic schemes (truncation, and selective drop-out of every second letter) that were run with and without the generation of a check letter, the check letter improved the performance.

For subject words and for proper names, the best

[13] Bourne, C. P., and D. F. Ford, "A Study of Methods for Systematically Abbreviating English Words and Names," *Journal of the Association for Computing Machinery*, Vol. 8, No. 4, pp. 538–552 (October 1961).

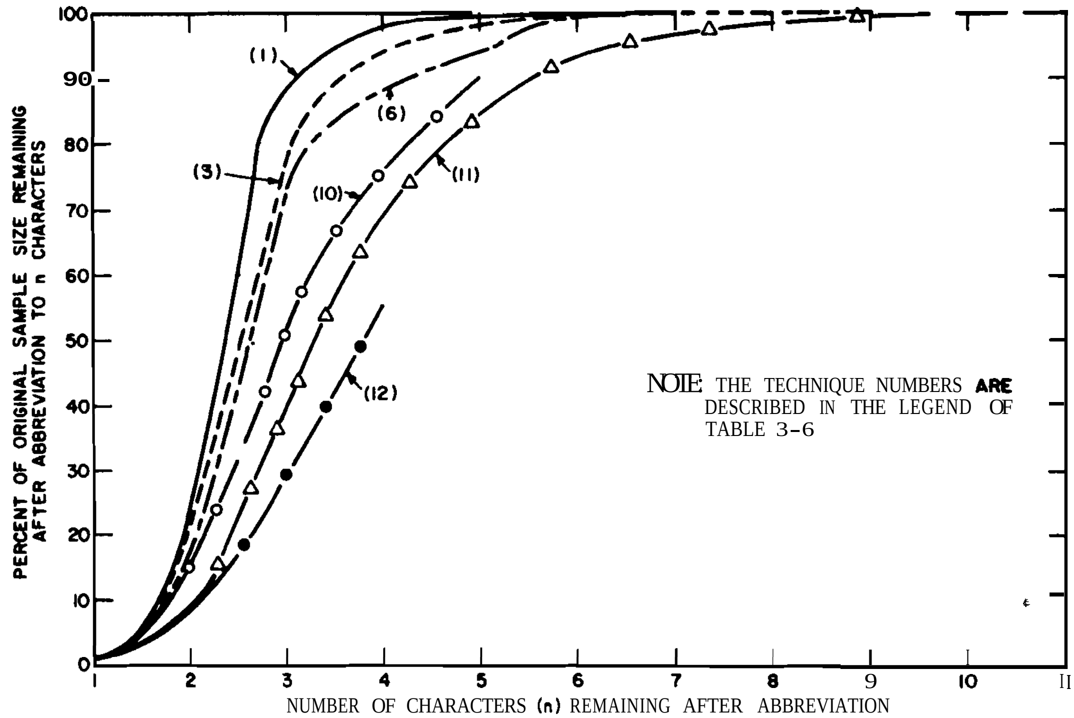


FIG. 3-6 Operating characteristics for the abbreviation schemes tested with subject words.

performance was obtained by dropping every other letter and generating a check letter, i.e., selective drop-out, n = 2, with a check letter.

For subject words, several of the performance curves crossed each other, so that no single scheme was "best," i.e., the "best" scheme varied with the number of letters allowed in the abbreviation. Because of the large number of tests on the subject words, the curves were plotted in two different figures, making it difficult to compare the techniques. Therefore, Table 3-6 was produced to show the rankings, by performance, of all the tested techniques, for abbreviated word sizes of 3, 4, 5, and 6 characters. The table shows that for subject words and an allowable field size of 3 to 6 characters, the best performance was obtained by dropping every other letter and generating a check letter. The test results for proper names did not intersect on the chart and thus there was a single "best" technique for all degrees of abbreviation. Acceptance of these conclusions should be tempered by recognition that not all of the described techniques have been tested.

The subject words of the data base ranged in length from 1 to 23 characters. However, with most of the abbreviation techniques, the data could still be represented by 8 to 10 characters while retaining the same degree of discrimination as the original words, i.e., the number of unique file items remained the same after abbreviation even though the field size had been restricted to 8 or 10 characters. The proper names of the data base ranged from 6 to 22 characters after editing. However, these names could also be reduced to 8 to 10 characters while retaining a high percentage of their original uniqueness. These results would probably vary with the total size and configuration of the data base being processed.

Assigned Coding Methods

Assigned coding methods require the use of detailed code tables for both the coding and decoding operations. Any type of material can be coded by this technique; single words or even large groups of words can be condensed into a concise table entry. For example, in 1874 Bloomer published a code dictionary of frequently used sentences and statements, to reduce the cost of cabling and to provide a cer-

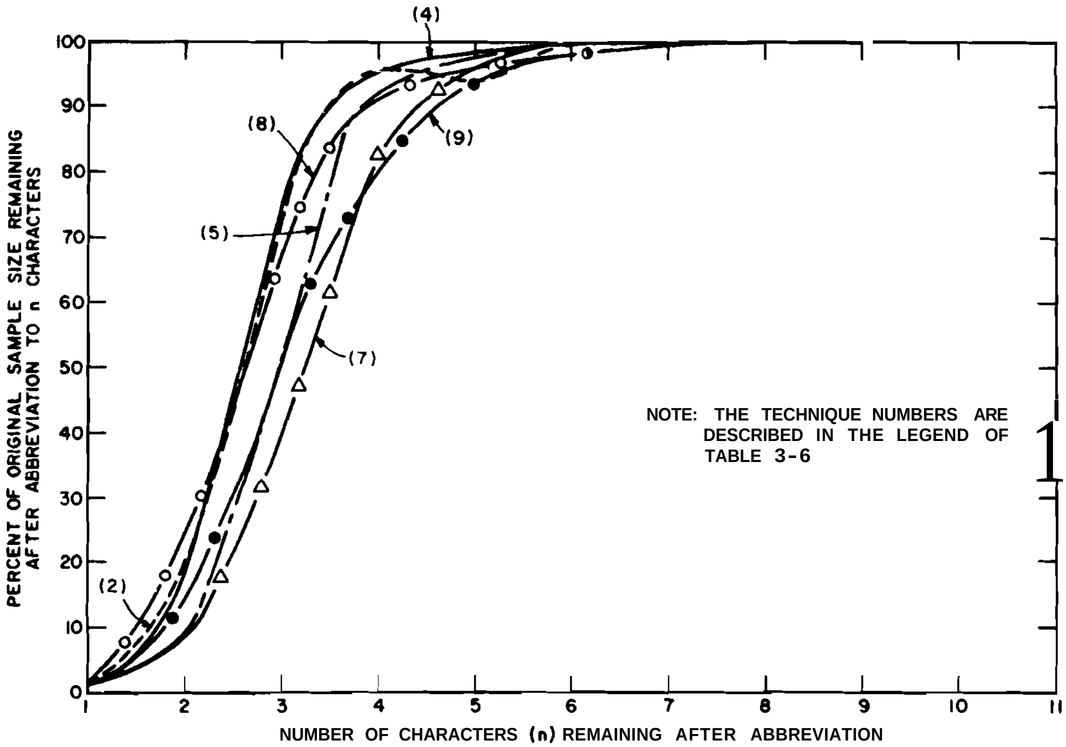


FIG. 3-7 Operating characteristics for the abbreviation schemer tested with **subject words.**

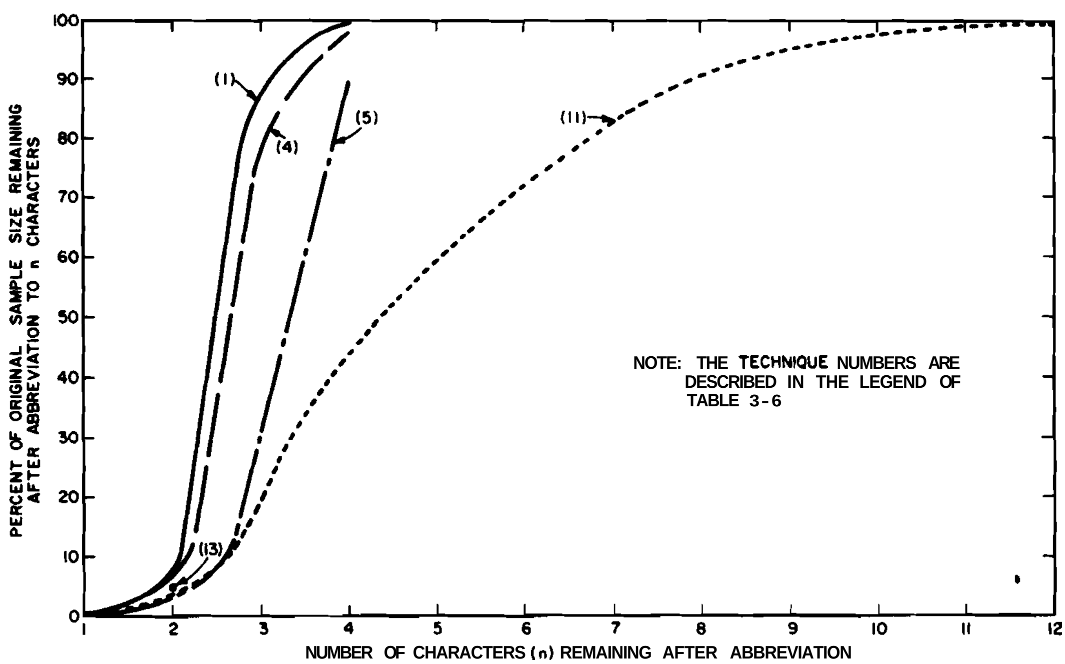


FIG. 3-8 Operating characteristics for the abbreviation schemer tested with proper **names.**

TABLE 3-6

Ranking of Techniques by Performance on Subject Words

	Allowable Field Size (No. of Characters) 3	4	5	6
Technique Number	1	1	1	1
	3	2	4	2
	4	4	5	4
	6	3	3	6
	2	5	7	3
	8	8	8	5
	9	6	9	7
	10	7	2	9
	5	9	6	8
	11	10	10	11
	7	11	11	
	12	12		

Note: The best techniques are at the top of the list.

Legend

Technique No.	Description of Technique
1	Selective drop-out (n = 2) with a check letter
2	Selective drop-out by separate rankings of character usage for each letter position
3	Selective drop-out by a single ranking of bigram usage
4	Selective drop-out (n = 2)
5	Selective drop-out ($n = 3$)
6	Selective drop-out by a single ranking of letter usage
7	Selective drop-out (n = 4)
8	Truncate right end and add check letter
9	Vowel elimination
10	Shuffle
11	Truncate right end
12	Truncate left end

tain degree of message security.[14] Examples of his codes are given below, with their assigned meaning:

Fient: Panic in the market. If you want to sell, telegraph immediately.

Phelan: Be careful in dealing with them, as they will take advantage, although their means are ample.

[14] Bloomer, J. G., *Bloomer's Commercial Cryptograph—A Teletype and Double Index-Holocryptic Cipher* (A. Roman & Co., San Francisco, California, 1874).

Hoop: Remit to us in gold coin by express, letting portions come daily until all is sent; limiting shipments to the amout of

A similar code dictionary was published in 1913,[15] which included the following:

Loranthus: Sell out at once; do not delay.
Coloring: Do not decide until you receive my letter.
Adjunct: Have had no letters from you for a fortnight or more. Are all well?
Reversible: We all unite in wishing you a very Merry Christmas.

Commercial, government, and amateur radio and teletype operators use a variety of standard code terms to save transmission time and effort. Examples of these code terms are:

GM	Good morning
CUL	See you later
BCNU	I'll be seeing you
73	Best regards
QRM	Are you being interfered with? I am interfered with.
QRU	Have you anything for me? I have nothing for you.

Some military organizations use standardized abbreviated writing for applications where data must be recorded or transmitted as quickly as possible.[16] An example of this type of notation is given by the following:

POW states he was a Msgr with the 22nd M Tk Bn of the 316th Fus Mecz Div. CO of the Div is Col Hans Heiner, a tough and energetic Off who has written several FMs on Armd tactics and Sup. Div was formed from the remnants of the old 41st Med Tk Brig under a new TOE effective 1 Jan 1954, and was a Repl unit until recently committed to the line. Elms of Svc units from the FIRST Fus Rif Army (Reinf) have been Atchd.

The message is expanded here to illustrate the degree of compression that was obtained.

The prisoner of war states that he was a messenger with the 22nd Medium Tank Battalion of the 316th Fusilier Mechanized Division. The commanding officer of the Division is Col. Hans Heiner, a tough and energetic officer who has written several field manuals on armored tactics and supply. The Division was formed from the remnants of the old 41st Medium Tank Brigade

[15] ——, *The Adams Cable Codex,* 10th ed. (F. O. Houghton & Co., Boston, 1913).

[16] *Military Symbols,* Dept. of the Army Field Manual FM 21-30 (June 1951).

under a new table of organization effective 1 January 1954, and was a replacement unit until recently committed to the line. Elements of service units from the 1st Fusilier Rifle Army (reinforced) have been attached.

This type of abbreviation can easily be generated and interpreted after practice. The message can be filed in abbreviated form and later expanded into the original text.

Some computer and data processing systems use special name-coding techniques to identify customers or accounts. In many cases these codes are combinations of fragments of the individual's name and address, sometimes with additional data such as the type of account or details of the account. One particular name-coding scheme derives an 11-character code from elements of the person's family name and initials, birthday, sex, and a check digit.[17] The first 6 characters are the first 4 letters of the family name followed by the first and middle initial. The next character represents the month of the birthday, and the next 2 characters represent the day of birth, The next character represents the unit digit of the year of birth, and the last character is a check character for the preceding 10 characters. Missing letters are replaced by an X.

Another name-coding scheme, which has been used successfully for driver licensing and motor vehicle registration, uses elements of the person's name, age, and physical characteristics to form a code 15 characters long.[18] It takes advantage of the fact that while there are about 1¼ million different names in the United States, approximately 50 percent of the people have only about 1500 different names (Smith, Jones, etc.). A table look-up operation identifies each of approximately 8000 of the most common names with a four-digit number. The use of the first letter of the name plus the four-digit number is said to produce an alphabetic sequence which is 85 to 95 percent accurate. The four-digit number is then appended to the first letter of the family name. To these 5 characters representing the last name, 3 more digits are appended representing the first four letters of the first name. Over 90 percent of first names can be found in approximately 770 four-letter combinations. Two digits are then appended to represent the middle initial. The remaining 5 digits of the code are based on personal data and characteristics; 1 character for month of birth, 2 digits for year of birth, 1 character for combination of sex and eye color; and 1 digit for range of height. This coding scheme reportedly produced code entries which were completely unique for several years of operation of the driver licensing and motor vehicle registration departments for a large state. Other examples of specific account coding techniques are referenced below.[19]

Many elaborate assigned coding schemes have been developed, especially by large publishers, for maintaining mailing lists having up to 40 million names. Business and industrial organizations also maintain large lists for mailing internal reports and notices, company publications, press releases, and mail advertising. As mentioned earlier, there are many organizations that specialize in collecting, maintaining, and selling lists of names for direct mail advertising, like the familiar '(Occupant" addresses. In many of these cases, the coding schemes are quite elegant, and are implemented with computer or punched card equipment. A well-organized company mailing list may often consist of 15 or 20 different lists, including lists for each company publication, for different types of meeting announcements, for various types of press releases, and for various groups of customers. The list might also include a geographic code showing the general area and major city of the addresses, and distinguishing between classes of direct and residue cities as specified by the Post Office. The files may also be arranged so that the lists are organized in a hierarchical order by geographic region, city, street, name, and number instead of alphabetically by addressee. For mailings, this arrangement can save the Post Office a tremendous amount of effort; some of the larger publishers are able to save the Post

[17] Luhn, H. P., *A Personal Identification Code,* IBM Corp., Report AM-1.01.022.001 of the Advanced Systems Development Division, Yorktown Heights, New York (August 19, 1959).

[18] IBM Corp., *General Information Manual: A Unique Computable Name Code for Alphabetic Account Numbering,* Report F20-8052 of the Data Processing Division, White Plains, New York (1960).

[19] Oettinger, A. G., "Account Identification for Automatic Data Processing," *Journal of the Association for Computing Machinery,* Vol. 4, No. 3, pp. 245–253 (July 1957).

Taunton, B. W., "Name Code—A Method of Filing Accounts Alphabetically on a Computer," *Data Processing,* Vol. 2, No. 3, pp. 23–24 (March 1960).

Trueman, R. E., "Development of an Efficient Account-Numbering Method," *Management Science,* Vol. 7, No. 3, pp. 265–279 (April 1961).

Office hundreds of thousands of dollars per year by organizing their mailing lists in this way. In return for this pre-sorting service, the Post Office allows reduced rates, and some saving in delivery time usually results. The user with a geographic file problem might consider the use of a numerical code for states, counties, and cities of the United States. A code system of this type, which also includes an indication of the cities' or counties' recent population total is available from IBM.[20] The same organization has also published some other useful general explanations of coding schemes.[21]

Another common example of assigned coding is the notation used in many libraries with the Universal Decimal, Library of Congress, Bliss, or other classification schemes. The coding or notation is used for convenience to represent the various subject classes. The following examples describe some of the notation used by several subject classification schemes:

Classification	Sample Notation	Term
Dewey Decimal	621.312 136	Winddriven electric plants
Library of Congress	TK 1001	General transmission of power
Universal Decimal	551.507.362.2	Artificial satellites

A distinction should be made between the classification scheme and its system of notation or coding. For any classification scheme, a variety of notations could be used, and each would have little effect on how well the scheme organized a collection of information. The notation of a scheme is a matter that should be considered independently of the basic question of how information should be grouped and organized. Unfortunately, classification systems are often criticized primarily on the basis of their notation, even though it is quite possible to completely change the system of notation without changing the basic structure of the classification scheme.

One further example of word coding is given by the ASTIA Thesaurus Code Manual, a table of code numbers for each descriptor used in their indexing system.[22] The use of fixed-length numeric code numbers instead of the full alphabetic spelling of the descriptors simplifies the subsequent machine handling by the data processing equipment. A seven-digit number is assigned to each descriptor so that numeric sequence and alphabetic sequence coincide. The machine codes for astronomy, beeswax, and cats are 0534000, 0711000, and 1098000, respectively.

Another good example of an assigned coding scheme is the use of an empirically derived table to code the names or subject headings into equal-interval groups. Such a scheme is useful, because in many practical applications it may be desirable to break a file or list of people's names into groups of equal size. This is easy to do if the file or list already exists. However, it may be necessary to make the division before the file items are collected, and thus before the characteristics of the file can be known. An example would be the establishment (by the letters of the name) of equal-size groups of new students to be registered during the eight time periods available for class registration. Another example would be the establishment of call numbers for library books to help group the books on the shelves within a particular subject classification in alphabetic order by the author's name. Such call numbers have been derived in many libraries by using equal-interval tables such as the ones described in this section. The tables most commonly used for the establishment of library call numbers have been the Cutter-Sanborn tables, first developed by C. A. Cutter in the 1880's and based on lists of author's names, and later modified by K. E. Sanborn.[23] Several other equal-interval tables have been developed for library call numbers.[24]

These situations can be handled by using empirical data like those in Table 3-7 which show the

[20] IBM Corp., *Numerical Code for States, Counties, and Cities of the United States*, Brochure X21-4653, IBM Corp., New York (1958).

[21] IBM Corp., *Account Numbering and Self-checking Number Systems*, Brochure G22-8502-0, IBM Corp., New York (June 1959).

Modern Coding Methods, Brochure X21-3793-6, IBM Corp., New York (undated).

Reference Manual: Indez Organization for Information Retrieval, Brochure C20-8062, IBM Corp., Data Processing Div., White Plains, New York (1961).

[22] *ASTIA Thesaurus Code Manual*, an un-numbered report of the Armed Services Technical Information Agency (June 1961).

[23] *C. A. Cutter's Alphabetic-Order Table—Consonants except S Altered and Fitted with Three Figures by Miss Kate E. Sanborn* (Library Bureau, Boston, 1896).

[24] W. C. Berwick Sayers, *A Manual of Classification for Librarians and Bibliographers*, 3rd ed., rev., Chapt. 23 (Grafton & Co., London, 1955).

TABLE 3-7

Numerical Breakdown for Coding Names into Equal-Size Groups

Aa – Alk	01	Ell – Eu	27	Lan–Led	52	Rip – Rog	76
All – Andq	02	Ev – Feq	28	Lee – Lewi	53	Roh – Rue	77
Andr – Aul	03	Fer – Flem	29	Lewj – Lol	54	Ruf – Sans	78
Aum – Ball	04	Flen – Fram	30	Lom – Lym	55	Sant – Schn	79
Balm – Bar	05	Fran – Gaa	31	Lyn – Magm	56	Scho – Sec	80
Bas – Belk	06	Gab –Gel	32	Magn – Marsg	57	Sed – Shel	81
Bell – Beu	07	Gem – Gla	33	Marsh – Mcat	58	Shem – Simo	82
Bev – Bln	08	Glb – Grac	34	Mcau – Mcem	59	Simp – Smith, G.	83
Blo – Bour	09	Grad – Grid	35	Mcen – Mcni	60	Smith, H. – Sot	84
Bous – Breec	10	Grie – Hag	36	Mcnj – Mey	61	Sou – Stans	85
Breed – Brown, H.	11	Hah – Ham	37	Mez – Mir	62	Stant – Stev	86
Brown, I. – Bun	12	Han – Harrisa	38	Mis – Mord	63	Stew – Stuf	87
Buo – Calc	13	Harrisb – Hay	39	More – Muk	64	Stug – Tag	88
Cald – Carq	14	Haz – Hero	40	Mul – Nea	65	Tah – Thei	89
Carr – Chanc	15	Herp – His	41	Neb – Niel	66	Thej – Tis	90
Chand – Claq	16	Hit – Hon	42	Nielm – Oha	67	Tit – Tuq	91
Clar – Cofe	17	Hoo – Hua	43	Ohb – Paf	68	Tur – Vaq	92
Coff – Con	18	Hub – Hya	44	Pag – Patte	69	Var – Walk	93
Coo – Coy	19	Hyb – Jam	45	Pattf – Pes	70	Wall – Wat	94
Coz – Culp	20	Jan – Johnson, R.	46	Pet – Pk	71	Wau – Wer	95
Culq – Dau	21	Johnson, S. – Kac	47	Pl – Prh	72	Wes – Wh	96
Dav – Dek	22	Kad – Kel	48	Pri – Rand	73	Wi – Wilm	97
Del – Dim	23	Kelm – King	49	Rane – Ren	74	Wiln – Wom	98
Din – Doy	24	Kinh – Kor	50	Reo – Rio	75	Won – Wyc	99
Doz – Dye	25	Kos – Lam	51			Wyd – Z	100
Dyf – Elk	26						

Source: This breakdown is a modification of the results of a detailed study of 63,000 names by the Records Bureau of Stanford University. To simplify the use of this table, some of the groups have been rounded off to simpler demarcation points.

statistical distributions of names that have been observed with other representative files. Table 3-7 is divided alphabetically into 100 groups, each group representing approximately 1/100 of the entire file. If, for example, a large and unknown group of people is to be divided into 10 groups, then the end of every tenth group in Table 3-7 could be used as a separation point.

PRIME NUMBER CODING

For coordinate indexing systems, a workable coding scheme can be developed in which each subject heading or Uniterm is represented by a unique prime number, and a particular document represented by the arithmetic product of all the pertinent prime numbers. In this way, the subject index of a collection would consist of a list of numbers, with one number for each file item. A search is made by arithmetically dividing the search primes, or the product of the search primes, into the number that describes a particular document. The search criteria are satisfied if there is no fractional remainder after division.

As an example, consider a system which had a code dictionary such as the following, in which each descriptor was represented by a unique prime number.

Descriptor	Code Number (Primes)
Abstracts	2
Aperture cards	3
Artificial language	5
Automata	7
Automatic abstracting	11
Automatic indexing	13
Automatic programing	17
Bibliographies	19
Bliss classification	23
Cataloging	29
Character recognition	31
	.

A document which was characterized by the three descriptors Automatic indexing, *Bliss* classification, and Cataloging, would be represented by the number $13 \times 23 \times 29 = 8671$. That is, this single number, 8671, would uniquely represent the joint occurrence of those three descriptors, since with the given table, there is no other way to derive the number 8671. Each item in the file would have its subject content represented by a single concise number. For searching purposes, each of these compound numbers would be divided by the code number for the search term, to find those numbers that could be divided without giving a finite remainder. For example, if our file item represented by 8671 were to be examined during a search for Automatic indexing, the search examination would consist of the division operation $8671 \div 13 = 667$, which produces an answer with no fractions. Aside from the three terms used for indexing (13, 23, 29) no other terms in the code dictionary can divide into 8671 without yielding a fraction remainder.

This particular scheme is awkward and cumbersome for the human user, and is not suited for manual systems. However, its numerical nature makes it amenable to mechanization by computers or other devices that can easily form the compound numbers and do the division for the search operations. Another difficulty with the system is that as the size of the dictionary or authority list grows, so does the length of the number that represents the average document. The magnitude of the final product depends upon the number of descriptors used and the size of the indexing vocabulary. If a dictionary of several thousand words is required, and if prime numbers are assigned in ascending order (2, 3, 5, . . .), then the last items in the vocabulary will be represented by large numbers, as shown in the following list.[25]

The	50th prime number is	227
	100th	523
	200th	1,217
	500th	3,559
	1,000th	7,907
	5,000th	48,593
	10,000th	104,723
	15,000th	163,819
	20,000th	224,729

If the dictionary has 10,000 descriptors, then a document with 12 descriptors would require a coding field sufficiently large to hold a number slightly less than $(104{,}723)^{12}$. The magnitude of this number is approximately 1.8×10^{60}, which can be represented by 61 decimal digits. A procedure of assigning the smaller prime numbers to the most frequent descriptors would reduce the average code length. In any case, the formation and division of these long numbers will pose some awkward problems for computing equipment as well as for people.

The first references to prime number coding in connection with document searching were made by Buck in 1958.[26] Buck made use of a mathematical coding problem that was solved earlier by Gödel, and referred to the coding as *Gödel* indices. Kokie later described an operating system for an IBM 650 computer that used prime numbers of up to 4 digits to index a file of documents.[27] There are approximately 1200 prime numbers with 4 digits or less, so that a reasonable dictionary size can be maintained for a small file of documents. Kokie's restriction to 4 descriptors per document was necessary because

[25] Lehmer, D. N., List of Prime Numbers from *1* to *10,*006,791, Publication No. 165, Carnegie Institute of Washington, D.C. (1914).

[26] Buck, R. C., Studies in *Information* Storage and Retrieval: I. On the Use of *Gödel* Indices in Coding, MRC Technical Summary Report No. 53, University of Wisconsin, Math Research Center, Madison, Wisconsin (October 1958), AD-207 325; and "Studies in Information Storage and Retrieval: On the Use of Gödel Indices in Coding," American Documentation, Vol. 12, No. 3, pp. 165–171 (July 1961).

[27] Kokie, J. E., "An Experimental Numerical Information Retrieval System," paper presented at the 15th National Conference of the Association for Computing Machinery, Milwaukee, Wisconsin (August 1960).

the computer could not conveniently handle the product of larger primes.

At this same time period, Cockayne and Hyde experimented with a prime number coding system for chemical compounds, using a computer searching system, and stated that they found this to be a way to achieve a savings in storage space over more conventional systems.[28] Some 25,000 organic chemical compounds were coded with prime numbers that represented their physical and other properties. This system was indexed by 208 different structural features, and used the first 208 primes. The smaller primes were allocated to the more frequently occurring features in an effort to keep the prime product to a relatively small number. For a given chemical compound, the primes corresponding to its structural features were selected and multiplied, and the resulting product was used as the code for that chemical compound. The test for the presence of a set of features consisted of dividing the code number by a factor, formed from the primes corresponding to the particular set, and testing for zero remainder. This process selected those compounds that had at least the specified features. It was found that even with the possibility of coding 17 features per chemical, the largest code number encountered, approximately 10^{22}, was very much less than the computed worst case, approximately 10^{52}.

In a later publication, Lamm suggested that prime number coding as previously described could be usefully modified into a positional scheme, in which the code for each element is a prime number selected in such a way that the digits occupying various positions have special meanings.[29]

SUPERIMPOSED CODING

Superimposed coding was first developed, analyzed, and used as a coding technique for information storage and retrieval systems—manual and mechanized—around 1947.[30] Its principal advantage is a reduction in storage space requirements for card, tape, or computer systems. Other advantages, for machine handling, are the built-in logic capabilities and the alignment of the descriptor fields. The primary disadvantages are the possibility of ambiguities, the occurrence of false drops,[31] the requirement of coding and encoding by table look-up, and most important—the lack of a single composite indexing number that is convenient for human interpretation, notation, and retention. The false drop problem is not a serious one since code design techniques allow the frequency of false drops to be designed to any desired level.

The usual coding method consists of determining, from a code dictionary, the particular pattern of marks assigned to each relevant descriptor, and then superimposing the patterns of all these descriptors to form one composite pattern to represent the document. These patterns could be represented in many forms, such as the punched holes in an edge-notched card or numbers recorded on magnetic tape. The marks all exist only in one of two possible states (i.e., a hole is either punched or unpunched, a number is either 0 or 1). This situation is referred to as *binary* coding, which is characterized by the fact that each code element can only exist in one of two

[28] Cockayne, A. H., and E. Hyde, "Prime Number Coding for Information Retrieval," Computer *Journal* (Great Britain), Vol. 3, No. 1, pp. 21–22 (April 1960).

Fairthorne, R. A., "Prime Number Coding for Information Retrieval" (a review of the article by Cockayne and Hyde), Computer *Journal* (Great Britain), Vol. 4, No. 1, p. 85 (April 1961).

[29] Lamm, E., "Prime-Number Coding," American Documentation, Vol. 12, No. 3, pp. 172–177 (July 1961).

[30] Isbell, A. F., "A Practical Application of a Punched-Card System Utilizing the Superposition of Codes," paper presented at the Division of Chemical Literature of the American Chemical Society, 114th National Meeting, Portland, Oregon (September 1948).

Mooers, C., *Application* of Random Codes to the Gathering of Statistical *Information,* Bulletin 31, Zator Co., Cambridge, Mass. (1949) (based on M.S. thesis, MIT, January 1948).

Mooers, C., "Zatocoding Applied to Mechanical Organization of Knowledge," American Documentation, Vol. 2, No. 1, pp. 20–32 (January 1951).

Mooers, C., "Zatocoding and Developments in Information Retrieval," Association of Special Libraries and Information Bureaux *Proceedings,* Vol. 8, No. 1, pp. 3–22 (February 1956).

Wise, C., "Mathematical Analysis of Coding Systems," Chapt. 21 in Punched Cards: Their *Application* to Science and *Industry,* 2nd ed., R. S. Casey et al., editors (Reinhold Publishing Corp., New York, 1951).

[31] "False drops" are defined as items that are located in response to a search request, but that have no relevance to the search question. They occur because the resulting indexing pattern on a particular document is not unique, and can be synthesized from several different combinations of descriptor code patterns.

states. The following illustration shows how the method might be used for the encoding of a document with 4 descriptors:

Descriptor	Code Pattern
Magnetic	010000000001100000
Computer	000000000010000011
Circuits	000000010001000010
Digital	000000000010000011
Resulting superimposed code pattern:	010000010011100011

Just as in the case of prime number coding, a single composite pattern is derived (with the aid of a code dictionary) to represent the subject indexing of that file item. However, in this case it is much more difficult to transcribe, read, or otherwise handle the composite pattern. Notice that this is really a code pattern, and not a specific number, although it could be interpreted as a binary number and converted to its equivalent in some other number base. As mentioned earlier, the pattern could be a set of punches in a card, a set of notches on the edge of a card, a set of binary numbers stored on magnetic tape, or a set of binary marks stored on a piece of microfilm. The search process consists of superimposing the patterns of the search descriptors to form a composite quiz pattern, then checking each file item to see if the quiz pattern is included in the pattern of the file item. A search is satisfied if there is a corresponding mark in the file item for every marked position in the quiz pattern.

The code patterns in the code dictionary are initially assigned to the descriptors on a random basis, primarily to distribute the marking uniformly over the whole coding pattern and enhance system performance. However, this is not essential to the operation of a superimposed coding system. One interesting modification of this random binary coding might be to use certain pairs of letters taken from the descriptor itself, rather than random numbers.[32]

To date, superimposed coding has been used primarily for edge-notched card systems. The complete card and coding system developed by Mooers (see Fig. 3-9) is the best example of such a system, and is referred to as Zatocoding.[33] In this figure, the reference card is notched with the codes for 7 descriptors. The code pattern for each descriptor has 4 marks (punches) in a total field of 40 positions. That is, in this example, there are 4 marks per descriptor in a field size of 40 coding sites. And instead of showing the binary pattern of marks for each descriptor (40 digits for each descriptor, consisting of 4 ones and 36 zeros), the card shows the *location numbers* of the positions on the card that are to be punched. Thus the descriptor *camera* is represented by 1, 8, 29, 34, meaning that those 4 positions are to be punched on the card. The top of this figure shows the similar action that takes place during the superimposition of the search terms. Note that the composite search pattern is "included" in the card's indexing pattern—hence the card would be selected.

Superimposed coding has also been used with a few tabulating card systems.[34] However, superimposed coding need not be restricted to these methods of implementation, and could be used to good advantage for computer searching systems or for other special equipment.[35]

Finding the Number of Different Code Possibilities with m Marks in F Positions

Many of the characteristics of superimposed codes can be described mathematically; some of these are discussed in more detail in later sections, using the following terminology:

[32] Gilbert, P. T., Jr., "An Optimal Punch Card Code for General Files," *American Documentation,* Vol. 9, No. 2, pp. 84–98 (April 1958).

[33] Brenner, C. W., "Experience in Setting Up and Using the Zatocoding System," Chapt. 11 in *Information Systems in Documentation,* Shera et al., editors (Interscience Publishers, New York, 1957).

Brenner, C. W., and C. N. Mooers, "A Case History of a Zatocoding Information Retrieval System," Chapt. 15 in *Punched Cards,* 2nd ed., R. S. Casey et al., editors (Reinhold Publishing Corp., New York, 1958).

Perry, J. W., "Superimposed Punching of Numerical Codes on Hand-Sorted Punch Cards," *American Documentation,* Vol. 2, No. 4, pp. 205–212 (October 1951).

[34] Schultz, C. K., "An Application of Random Codes for Literature Searching," Chapt. 10 in *Punched Carols,* 2nd ed., R. S. Casey et al., editors (Reinhold Publishing Corp., New York, 1958).

Luhn, H. P., "Superimposed Coding with the Aid of Randomizing Squares for Use in Mechanical Information Searching Systems," Chapt. 23 in *Punched Cards,* 2nd ed., R. S. Casey et al., editors (Reinhold Publishing Corp., New York, 1958).

[35] Mooers, C. N., *The Application of Simple Pattern Inclusion Selection to Large-Scale Information Retrieval Systems,* Technical Bulletin No. 131, Zator Co., Cambridge, Mass. (April 1959), AD-215 434.

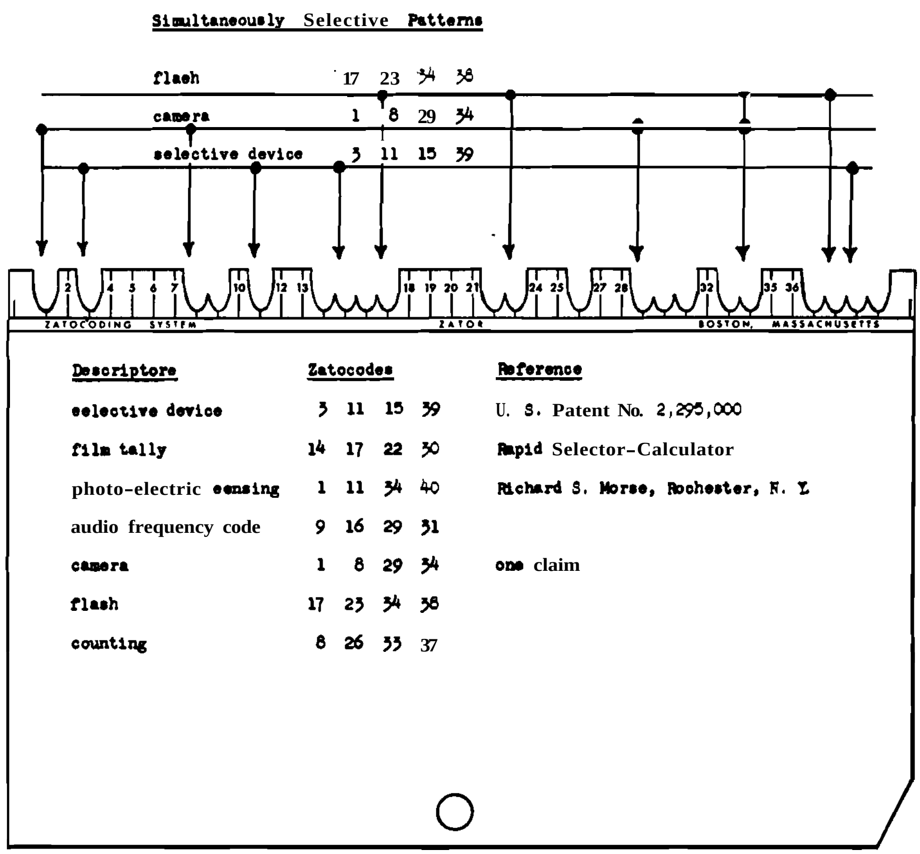


FIG. 3-9 Zatocoding, illustrated with a 5-by-&inch edge-notched Zatocard.

	← F code positions →	
	00100001001001 . . . 001	
	.	*d* descriptors with *m ones* or marks per descriptor
	00001000001000 . . . 100	
	00010000000101 . . . 100	
Superimposed pattern:	00111001001101 . . . 101	M *ones* or marks in the composite field

Using this terminology, Fig. 3-9 could be characterized as having 40 code positions $(F = 40)$, seven descriptors $(d = 7)$ with 4 marks per descriptor $(m = 4)$, and 21 marks in the composite field $(M = 21)$.

Nearly all superimposed code systems represent each term or descriptor by a fixed number of marks in a coding field of fixed size. For example, edge-punched card systems normally use one full side of the card (about 40 holes or code positions) for the coding field, with a random 4-mark pattern for each descriptor. However, there are a limited number of code possibilities for m marks in a field of F positions. For a given value of F (e.g., 40 for certain edge-punched cardsj, the value of m should be chosen to ensure that there are at least as many different code possibilities as there are descriptors in the indexing vocabulary. The actual number (N) of different code possibilities for m marks in a single field of F positions is given by simple combinatorial analysis as the binomial coefficient, expressed below in some of its common forms.[36]

$$N = \binom{F}{m} = \frac{F!}{m!(F-m)!}$$

[36] Feller, W., ***An Introduction to Probability Theory and Its Applications***, Vol. 1, 2nd ed. (John Wiley and Sons, New York, 1959).

Parzen, E., Modern ***Probability Theory and Its Applications*** (John Wiley and Sons, New York, 1960).

where

$$F! = F(F-1)(F-2)\ldots 1$$

and

$$0! = 1$$

This means, for example, that there are *10* **different** ways to mark *2* **holes** in a field with *5* hole positions. That **is,**

$$N = \binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5\cdot4\cdot3\cdot2\cdot1}{(2\cdot1)(3\cdot2\cdot1)} = 10$$

Using this formula, the number of unique code possibilities was computed for several values of m and *F* with the results shown in Table *3-8.* When the field size is an even number, the maximum number of unique combinations is obtained when m, the number of marks per descriptor, is one-half of the field size (i.e., $m = F/2$). If the field size is an odd number, the maximum number of combinations is obtained when the number of marks per descriptor is $\frac{1}{2}(F \pm 1)$.

The actual choice of *F* and m is usually a compromise among several conflicting objectives. It is convenient, especially for manual systems, to use a small value of m, to simplify the marking and punching. However, a large enough value of m must be used to provide the necessary number of code combinations. It will be shown later that the values of *F* and m are usually chosen in a systematic way to ensure an upper bound on the number of false drops that will occur during searching.

Number of *Marks* in the Superimposed Field

If each descriptor uses m marks in a single field, and d descriptors are superimposed, then the total number of marks, M, that actually appear in the final pattern will lie in the range of m and md. Estimates of the mean or average number of marked positions have been made by Mooers and Wise. One estimate suggests that the average number of marked positions will be given approximately by [37]

$$M \approx F - F\left(\frac{F-1}{F}\right)^d$$

Another estimate suggests that the average number of marked positions will be given approximately by [38]

$$M \approx F(1 - e^{-md/F})$$

37 **Wise, previous citation in Casey, p. 448 (fn. 30).**
38 **Mooers, C. N., *The Exact Distribution of the Number of Positions Marked in a Zatocoding Field,* Bulletin No. 73, Zator Co., Cambridge, Mass. (1952).**

TABLE 3-8

Number of Unique Code Combinations (*N*) for *m* Marks in F Positions

		Number of Marks per Descriptor (m)									
		1	2	3	4	5	6	7	8	9	10
	1	1									
	2	2	1								
	3	3	3	1							
	4	4	6	4	1						
Field	5	5	10	10	5	1					
Size (F)	6	6	15	20	15	6	1				
	7	7	21	35	35	21	7	1			
	8	8	28	56	70	56	28	8	1		
	9	9	36	84	126	126	84	36	9	1	
	10	10	45	120	210	252	210	120	45	10	1
	15	15	105	455	1,365	3,003	5,005	6,435	6,435	5,005	3,003
	20	20	190	1,140	4,845	15,504	38,760	77,520	125,970	167,960	184,756
	40	40	780	9,880	91,390	658,008		—	—	—	—
	50	50	1,225	19,600	230,300	2,118,760	—	—	—	—	—

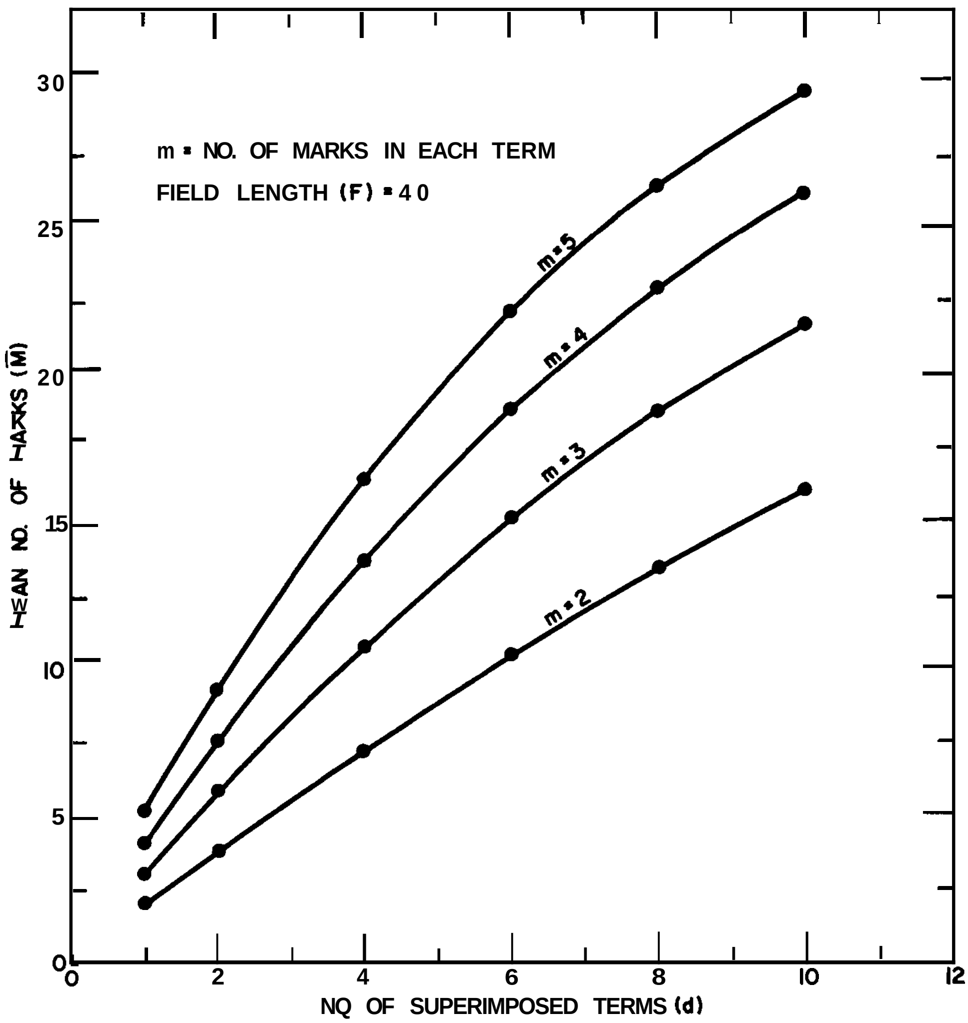


FIG. 3-10 Mean value ($\bar{M}$) of the distribution of marks in a field of 40 places resulting from the superimposition of *d* terms, each with m randomly placed marks.

Both of these expressions are only estimates. However, various approaches for obtaining the exact solution for the complete distribution of marked positions as well as for the mean value have been suggested and formulated explicitly.[39] There are some differences in the mathematical assumptions used by the different investigators—hence the results, and the computation procedures are somewhat different. The exact solutions are rather complicated functions, and for the sake of brevity are not repeated here. It is important to remember that these solutions are the exact solutions to a model that, unfortunately, does not exactly represent the true operational situation. To date, all of the models assume random, uniformly used descriptors, with no correlation between descriptors. All of these models differ from the real-life situations, in which some descriptors are used more thap others, and in which the descriptors are not completely independent of each other—for indexing or for searching. Although no detailed study has been made to show how well these models approximate the real situations, preliminary studies indicate that they

[39] Goldberg, J., et al., *Multiple Instantaneous Response Fib*, Final Report RADC-TR-61-233, Stanford Research Institute, Menlo Park, California (August 1961), AD-266 169.

Orosz, G., and L. Takács, "Some Probability Problems Concerning the Marking of Codes into the Superimposition Field," *Journal of Documentation*, Vol. 12, No. 4, pp. 231–234 (December 1956).

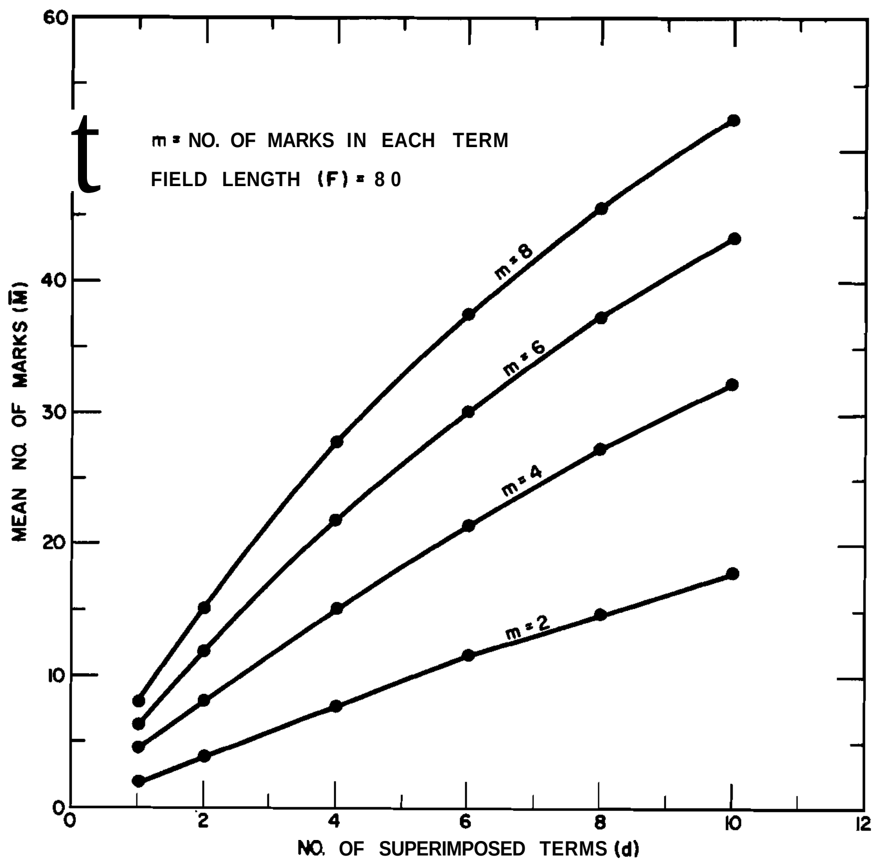


FIG. 3-11 Mean value ($\bar{M}$) of the distribution of marks in a field of 80 places resulting from the superimposition of d terms, each with m randomly placed marks.

provide a good first approximation. A few examples of the exact solutions of the mean number of marks, and their distributions in a coding field are given in Figs. 3-10, 3-11, and 3-12.[40]

Risk of False Drops

There is a finite possibility that the superimposition of several descriptors could yield a pattern that included the code pattern of one or more other descriptors. In this way, the final code pattern may appear to include irrelevant code terms. There is no way to look at the final code pattern in all instances and say with certainty exactly which descriptors were superimposed. The descriptors synthesized by the superimposition operation may result in an irrelevant response during a search. The possibility of this false drop is influenced by the coding and searching parameters, and can with proper design, be made as small as desired. The probability that a random interrogation will result in a false item being selected is often expressed as a "dropping fraction." One estimate of the upper bound to the probability of selecting an unwanted file entry is[41]

$$F_d \approx \left(\frac{M}{F}\right)^m$$

It has recently been shown that this upper bound is a good approximation which overestimates the exact random selection probabilities—thus it could be used as a conservative estimate to demonstrate the plausibility and general behavior of the code, even though it may be rather conservative for actual design.[42] The following approximation for the

[40] Goldberg report cited previously in this chapter, fn. 39.

[41] Mooers, C., *Zatocoding for Punched Cards,* Bulletin No. 30 of the Zator Co., Cambridge, Mass. (1950).

[42] Goldberg report cited earlier, fn. 39.

dropping fraction was derived from a different approach from that used for the previous approximation:[43]

$$F_d \approx \frac{\binom{M}{m}}{\binom{F}{m}} = \frac{\dfrac{M!}{m!(M-m)!}}{\dfrac{F!}{m!(F-m)!}}$$

An exact, and rather involved expression for the dropping fraction has been derived by Singleton, who has also computed the fractions for various code combinations.[44] The results of these computations indicated that the Wise approximation is too optimistic—frequently by several orders of magnitude—primarily because the derivation used the mean number of marks in the field instead of the exact distribution. The reader is cautioned that computed tables of dropping fractions that have used this expression, or variations of it which consider only the mean number of marks in a field, are probably too optimistic for serious code design work.[45] The exact computations for the probability of random selection for several different field sizes are given in Figs. 3-13, 3-14, and 3-15. These calculations appear to support Mooers' and Wise's rule that for single-field coding, a code field is best used when one-half of the places are marked.

[43] Wise, C. S., "Mathematical Analysis of Coding Systems," Chapt. 21, p. 457 in *Punched Cards,* 2nd ed., R. S. Casey et al., editors (Reinhold Publishing Corp., New York, 1958).

[44] Goldberg report cited earlier in this chapter, fn. 39.

[45] Taube, M., A. Kreithan, and L. B. Heilprin, "Superimposed Coding for Data Storage with an Appendix of Dropping Fraction Tables," Chapt. 5 in ***The Mechanization of Data Retrieval,*** Vol. 4 of the Studies in Coordinate Indexing Series (Documentation, Inc., Washington, D.C., 1957).

Stiassny, S., "Mathematical Analysis of Various Superimposed Coding Methods," ***American Documentation,*** Vol. 11, No. 2, pp. 155–169 (April 1960).

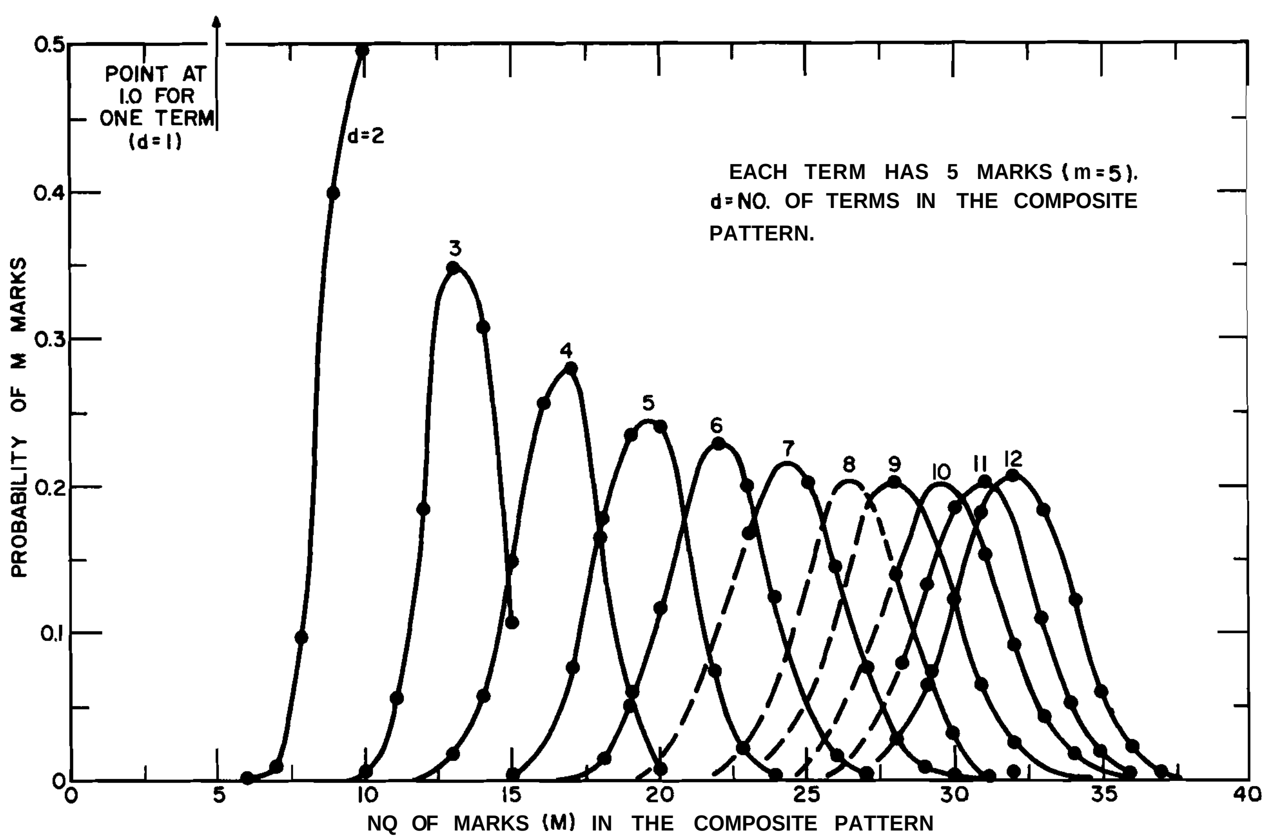


FIG. 3-12 Distribution of marks in a field of 40 placer from ruperimpositionr of *d* dexriptorr.

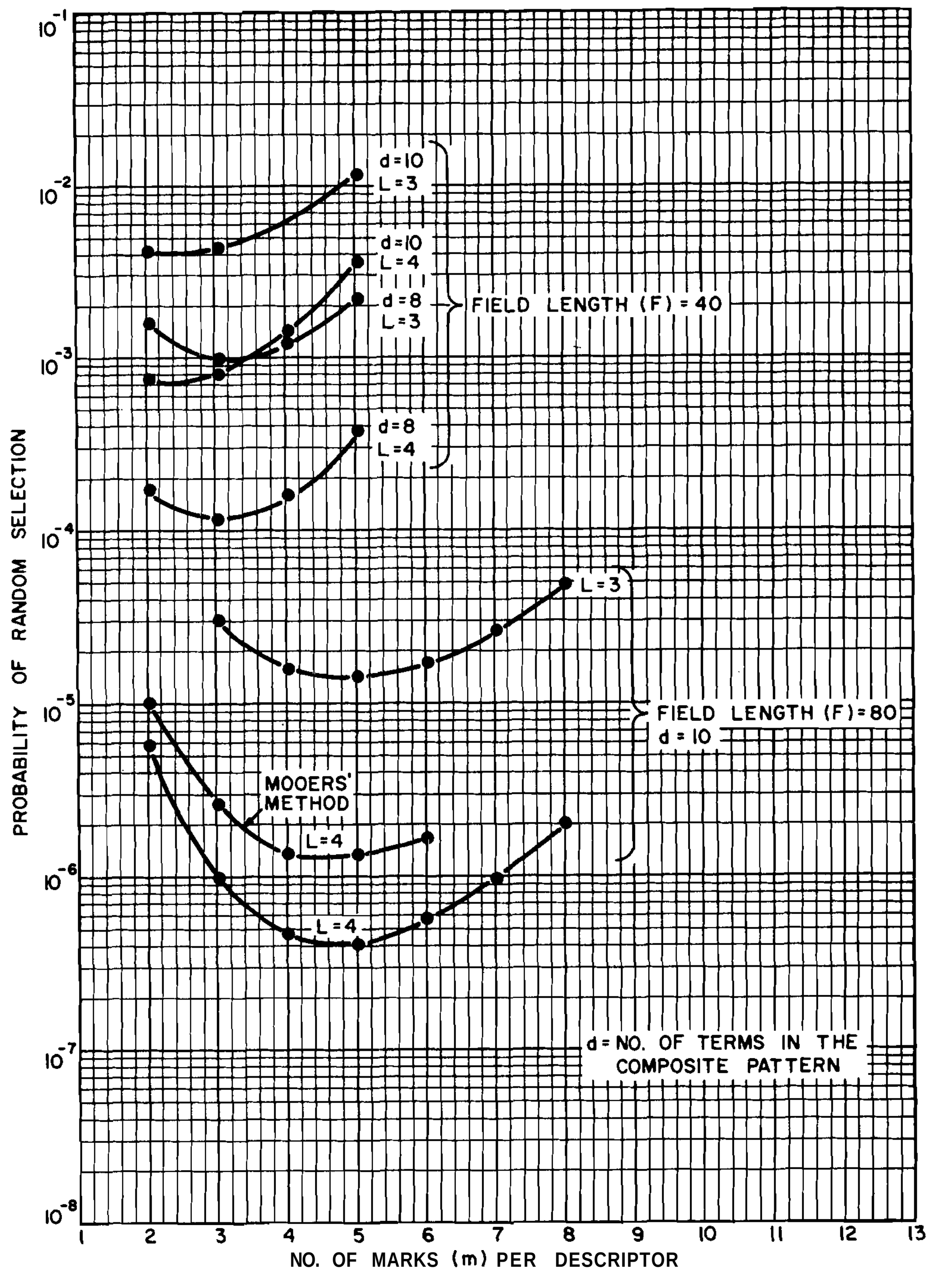


FIG. 3-13 Probability of random selection of a file item composed of d descriptors by a quiz with L descriptors, for various numbers of marks in basic descriptors.

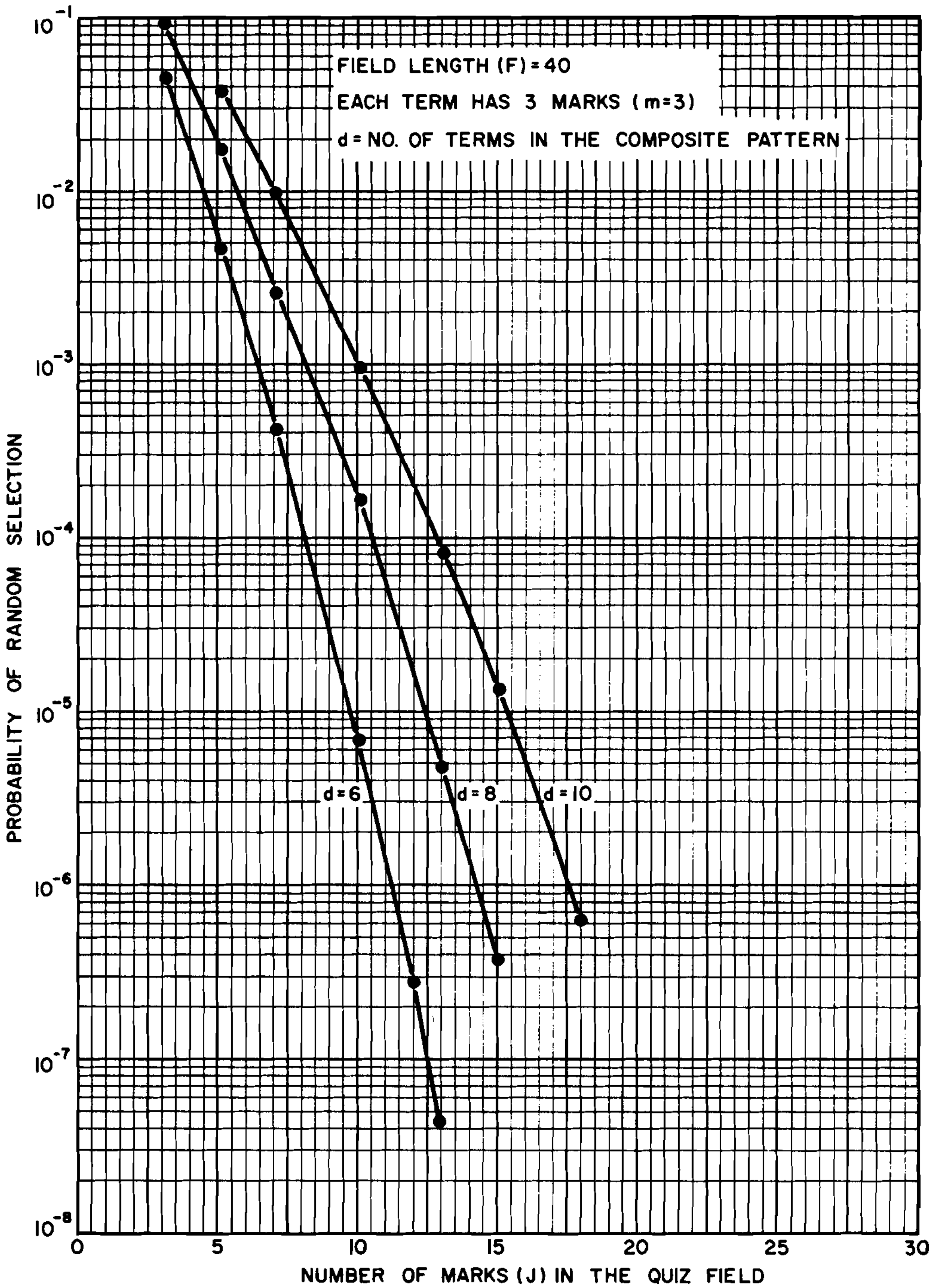


FIG. 3-14 Probability of random selection of a file item composed of d descriptors by a quiz with J marks.

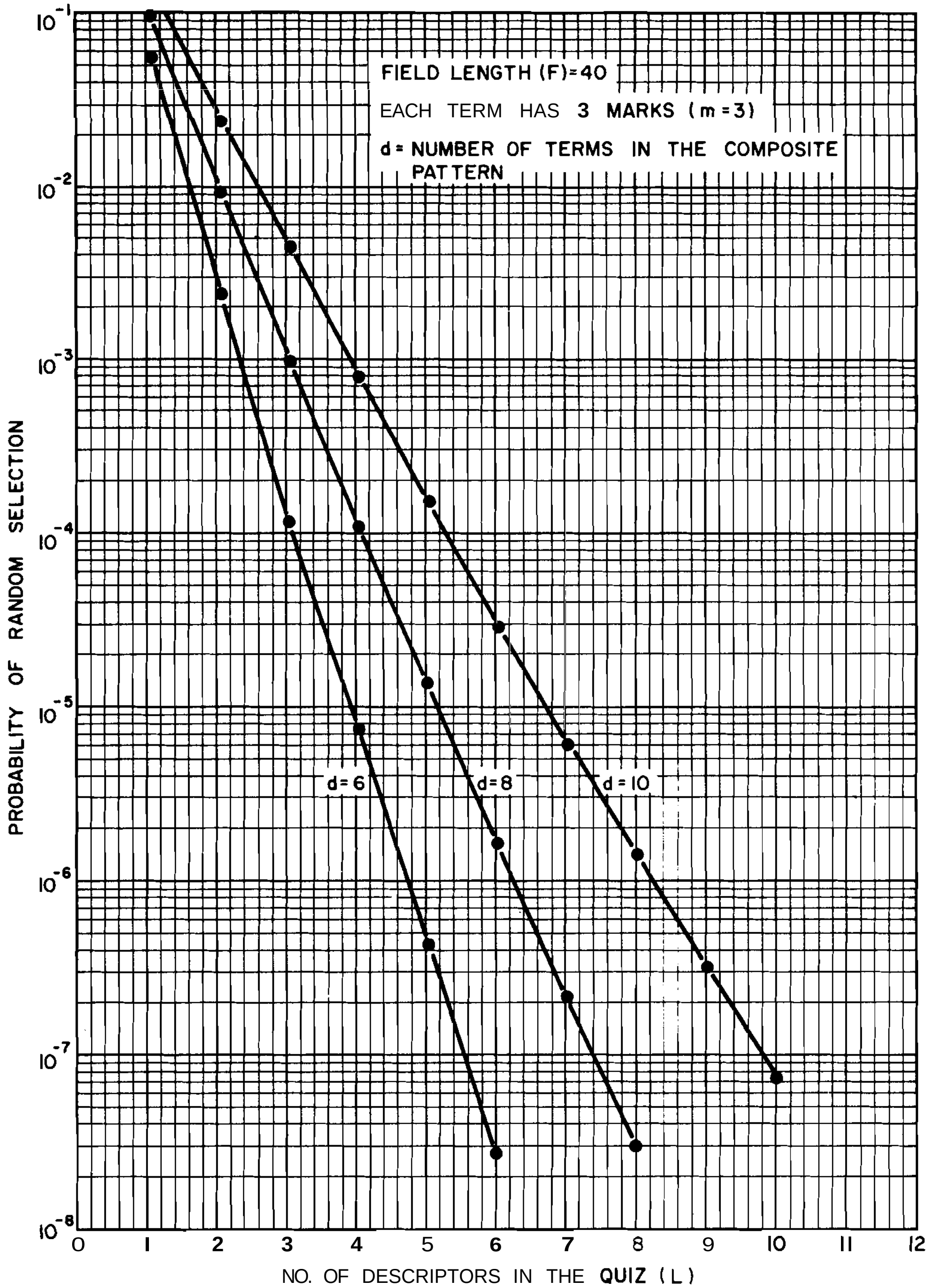


FIG. 3-15 Probability of random selection of a file item composed of d descriptors by a quiz with L descriptors.

How to Design the Superimposed Code

For relatively small retrieval systems, the user can generally adapt the systems and code parameters found to be successful by other users. However, for newer retrieval systems that require high performance of the superimposed coding system, a special study and code design may be in order. The design procedure is relatively simple, and considers the following parameters: [46]

- C the number of items in the total collection
- L the anticipated lower bound of the number of descriptors normally used for searching
- M the anticipated upper bound of the number of descriptors normally used for indexing
- R the tolerable noise ratio $= E_{max}/C$
- E_{max} the maximum number of false drops with L search descriptors
- F the length of the single fixed field for the superimposed code

In terms of these parameters, each descriptor code pattern should contain m marks (or binary ones), where

$$m = \left\langle \left(\frac{1}{L}\right)(-\log_2 R) \right\rangle$$

$$= \left\langle \left(\frac{1}{L}\right)(3.31)(-\log_{10} R) \right\rangle$$

where the symbols $\langle\ \rangle$ mean that the nearest integral value is to be taken. The least number of sites (F) that must be used to contain M descriptors is

$$F = \langle 1.445mM \rangle$$

For a sample calculation, assume the following parameters:

File size *(C)* = one million items

Minimum number of descriptors used for searching (L) = 3

Maximum number of descriptors used to index each item *(M)* = *12*

Maximum number of false drops tolerable with L search descriptors (E_{max}) = *100*

[46] Mooers, C. N., ***The Application of Simple Pattern Inclusion Selection to Large-Scale Information Retrieval Systems,*** Technical Bulletin No. *131,* Zator Co., Cambridge, Mass. (April 1959), *AD-215* 434.

Tolerable noise ratio (R) $= E_{max}/C$

$$R = \frac{E_{max}}{C} = \frac{100}{10^6} \equiv 10^{-4}$$

$$m = \left\langle \left(\frac{1}{L}\right)(3.31)(-\log_{10} 10^{-4}) \right\rangle$$

$$= \left\langle \frac{3.31}{3}(-1)(\log_{10} 10^{-4}) \right\rangle$$

$$= \left\langle \frac{3.31}{3}(-1)(-4) \right\rangle$$

$$= \langle 4.41 \rangle$$

$$= 4$$

and

$$F = \langle 1.445\ (4)\ (12) \rangle$$

$$= \langle 69.36 \rangle$$

$$= 69$$

Repeating this computation procedure for several different values of M, while keeping the same values of C, L, and E_{max} for this example, gives the following results:

Max. No. of Descriptors Used to Index Each Item *(M)*	Required No. of Marks per Descriptor *(m)*	No. of Code Positions Required *(F)*
3	4	14
6	4	29
12	4	69
20	4	96
40	4	191

The size of the coding field required, F, also varies with the size of the file. This slight variation is shown in Fig. **3-16,** which illustrates the degree to which the specification for E_{max} influences the size of coding field required.

Additional References

Barnard, G. A., "Statistical Calculation of Word Entropies for Four Western Languages," ***Institute of Radio Engineers Transactions of Professional Group on Information Theory,*** Vol. *1,* No. *1,* pp. *49-53* (March *1955).*

Barrett, J. A., and M. Grems, "Abbreviating Words Systematically," ***Communications of the Association for Computing Machinery,*** Vol. *3,* No. *5,* pp. *323-324* (May *1960).*

Bemer, R. W., "Do It By the Number—Digital Shorthand," ***Communications of the Association for Computing Machinery,*** Vol. *3,* No. *10,* pp. 530–536 (October *1960).*

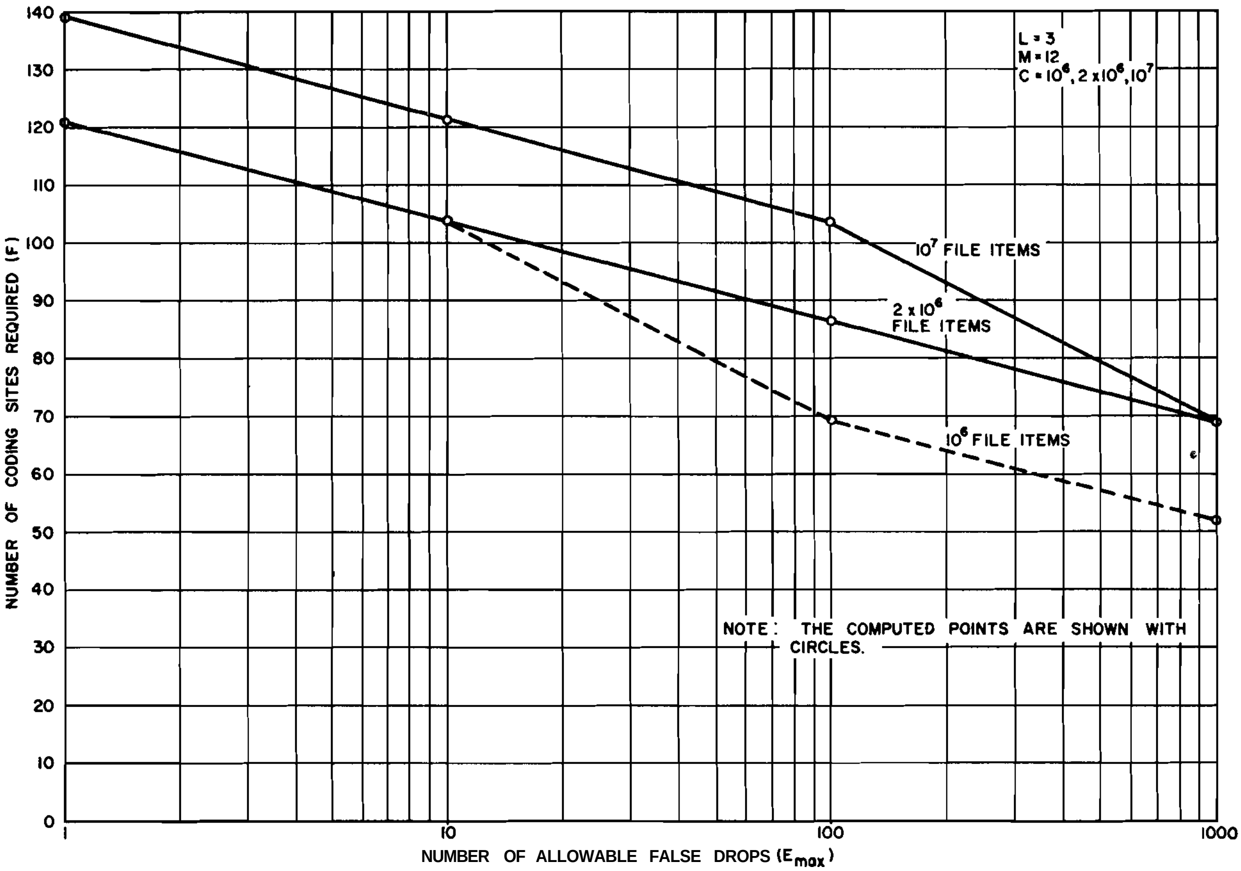


FIG. 3-16 Superimposed coding: memory requirements as a function of the number of allowable false drops.

Blair, C. R., "A Program for Correcting Spelling Errors," *Information and Control,* Vol. 3, No. 1, pp. 60–67 (March 1960).

Burton, N. G., and J. C. R. Licklider, "Long-Range Constraints in the Statistical Structure of Printed English," *American Journal of Psychology,* Vol. *68,* No. 4, pp. 650–653 (December 1955).

Carson, D. H., "Letter Constraints Within Words in Printed English," *Kybernetik,* Vol. 1, No. 1, pp. 46-54 (January 1961).

Chapanis, A., "The Reconstruction of Abbreviated Printed Messages," *Journal of Experimental Psychology,* Vol. *48,* No. 6, pp. 496–510 (1954).

Davidson, L., "Retrieval of Misspelled Names in an Airlines Passenger Record System," *Communications of the Association for Computing Machinery,* Vol. 5, No. 3, pp. 169–171 (March 1962).

Evans, M. W., C. K. McElwain, and F. Van Hoosen, *Machine Correction of Garbled English Text,* Massachusetts Institute of Technology Lincoln Laboratory Report (1960), AD-237 700.

Foster, M. B., "A Method of Representation, Storage and Retrieval of 13 Random Codes in a 4-Digit Number or 16 Random Codes in a 5-Digit Number," *Communications of the Association for Computing Machinery,* Vol. 5, No. 3, p. 165 (March 1962).

Frumkina, R. M., "Some Procedural Problems in Compiling Frequency Dictionaries (Statistical Structure of Dictionary and Text)," a translation of an article which appeared in the Russian-language periodical *Mashinnye Pereved i Priklad. Lingvistika* (Machine Translation and Applied Linguistics), No. 2(9), Moscow (1959). Translation available from Office of Technical Services, U.S. Dept. of Commerce Document No. JPRS:3599 (1960).

Garfield, E., "Information Theory and Other Quantitative Factors in Code Design for Document Card Systems," *Journal of Chemical Documentation,* Vol. 1, No. 1, pp. 70-75 (January 1961).

Korolev, L. N., "Coding and Code Compression," *Journal of the Association for Computing Machinery,* Vol. 5, No. 4, pp. 328-330 (October 1958).

Mandelbrot, B., "Simple Games of Strategy Occurring in Communication Through Natural Languages," *Institute of Radio Engineers Transactions on Information Theory,* Vol. PGIT-3, pp. 124–137 (March 1954).

Miller, G. A., "Some Effects of Intermittent Silence," *American Journal of Psychology,* Vol. 70, No. 2, pp. 311313 (June 1957).

Miller, G. A., and E. A. Friedman, "The Reconstruction of Mutilated English Texts," *Information and Control,* Vol. 1, No. 1, pp. 38–55 (September 1957).

Miller, G. A., and E. B. **Newman,** "Teat9 of a Statistical Explanation of the Rank-Frequency Relation for Words in Written **English,"** ***American*** Journal of Psychology, Vol. 71, pp. **209–218** (March **1958).**

Miller, G. A., E. B. **Newman,** and E. A. **Friedman, "Length-**Frequency Statistics for Written English," ***Information*** and Control Vol. 1, No. 4, pp. **370–389** (December **1958).**

Mooers, C. N., "Improvements in or Relating to Encoding Information on Punched Record Cards and Other Equivalent Media," Eritish Patent No. 681,902, **issued** 29 October 1952.

Mooers, C. N., "Manipulating Intelligence with a Random Code System," Canadian Patent No. **534,926,** issued 25 December 1956.

Newman, E. B., "The Pattern of Vowels and Consonants in Various Languages," American Journal of Psychology, Vol. 64, No. 3, pp. 369-379 (July 1951).

Newman, E. B., and L. **S.** Gerstman, "A New Method for Analyzing Printed English," Journal of ***Experimental*** Psychology, Vol. 44, No. 2, pp. **114–125** (August 1952).

Newman, E. B., and N. C. **Waugh,** "The Redundancy of Texts in Three Languages," Information and Control, Vol. 3, No. 2, pp. **141–153** (June **1960).**

Richmond, P. A., "A Simple Mnemonic **Large-Base** Number System for Notational Purposes," Journal of Documentation, Vol. 14, No. 4, pp. 208–211 (December 1958).

Shannon, C. E., "Prediction and Entropy of Printed **English,"** Bell System Technical Journal, Vol. 30, No. 1, pp. 50–64 (January 1951).

West, M., A General Service List of English Words with Semantic Frequencies **(Longmans,** Green, New York, **1953).**

Yngve, V. H., "Gap Analysis and Syntax," Institute of Radio Engineers Transactions on Information ***Theory,*** Vol. IT-2, No. 3, pp. 106–112 (September **1956).**

Punched Tabulating Cards
Paper Tape
Magnetic Tape
Machine Representation of Natural Languages

4

Machine-language representation

For machine retrieval or processing, some numeric, alphabetic, or special symbols must eventually be transformed to some physical representation—such as holes in a card or magnetized spots on a magnetic tape—in order that they can be stored and interpreted by the equipment. Most of the equipment in use today has a limited capability for representing and working with symbolic or textual information. Most of the methods for representing information on the various storage media (punched cards, punched tape, magnetic tape, etc.) were developed and adapted by the manufacturers of equipment that utilized these media. For example, both IBM and Remington Rand had developed their own separate card and code designs to go with their equipment; the cards are not compatible, and each of the coding systems uses only a small portion of the coding potential of the cards. There have been some attempts to develop a common or standard coding system for mechanized information processing equipment, but no definite standards have yet been established. The following sections describe the storage medium and coding systems used for punched tabulating card (tab card), paper tape, and magnetic tape systems.

PUNCHED TABULATING CARDS

The Storage Medium

IBM or Remington Rand punched tabulating cards are used for virtually all of the mechanized card-handling systems in the United States, with IBM supplying most of the installations.[1] Each of these card types (see Figs. 4-1 and 4-2) comes in a variety of sizes (normally 51-, 60-, 66-, and 80-column cards for IBM; and 32-, 40-, 54-, 66-, and 90-column cards for Remington Rand), but most organizations use the largest size possible. The smaller sizes are useful for some special applications, as will be mentioned later. The cards are available in a variety of colors, and are usually preprinted for the user's particular requirement. The characters (numbers, alphabetic symbols, and some special symbols), which are used as the restricted languages of the various Remington Rand and IBM punched card machines, are shown in Figs. 4-1 and 4-2.

The IBM card is most commonly used in such a way that each of the vertical columns on a card represents a single character. Consequently, the largest card size (80 columns) will permit 80 characters of information to be stored on a card. The Remington Rand card normally allows a maximum of 90 characters to be stored on a single card. The Remington Rand equipment normally operates with a smaller alphabet of characters than the IBM equipment. That is, it recognizes a smaller number of different symbols. Generally, the card equipment will accept and interpret the cards only in the manner of coding shown in Figs. 4-1 and 4-2. Arbitrary coding schemes devised by the card users

[1] Throughout this book, the IBM and Remington Rand type of card will often be referred to as a tab (tabulating) card, an expression in common usage.

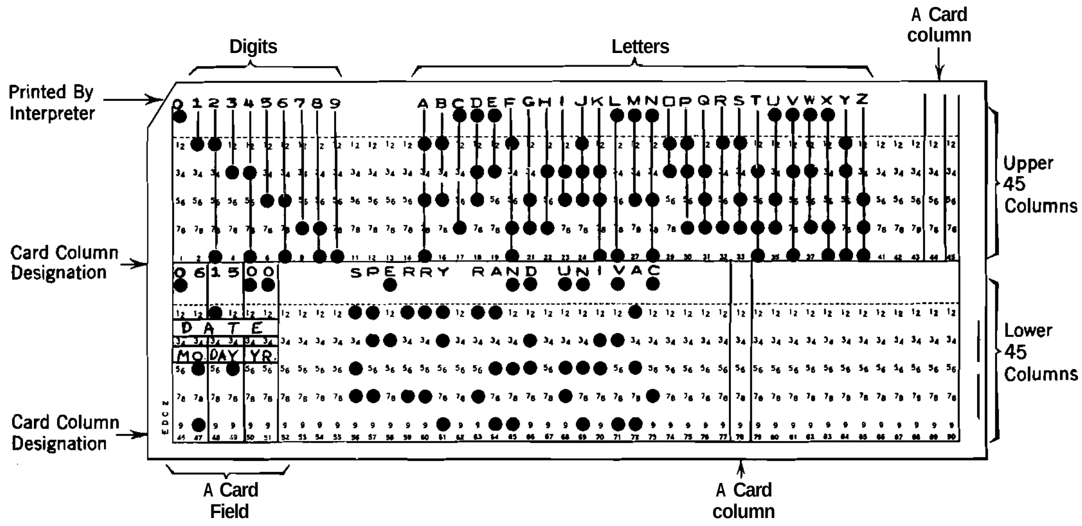


FIG. 4-1 Remington Rand 90-column card.

will usually be meaningless to the card-handling equipment.

From a practical standpoint the cards can tolerate a certain degree of rough handling, but do not work well when they have been folded, creased, warped, or stapled. Stapled or mutilated cards may jam the machines, or result in a misinterpretation of the data punched in them. Special card-editing equipment is available for the primary purpose of reading cards which have been handled by the public, in order to extract the illegible or manhandled cards. Another problem is that the cards gradually become frayed or damaged during repeated machine handling and may occasionally have to be regenerated. Some. card-handling problems may be aggravated by temperature or humidity effects; consequently, the cards should be stored in a cool, dry atmosphere.

Some card stock is available which is perforated or scored, to permit a portion of the card to be detached for subsequent card-handling operations. This detached portion might be used, for example, as a return stub for billing purposes, with a per-

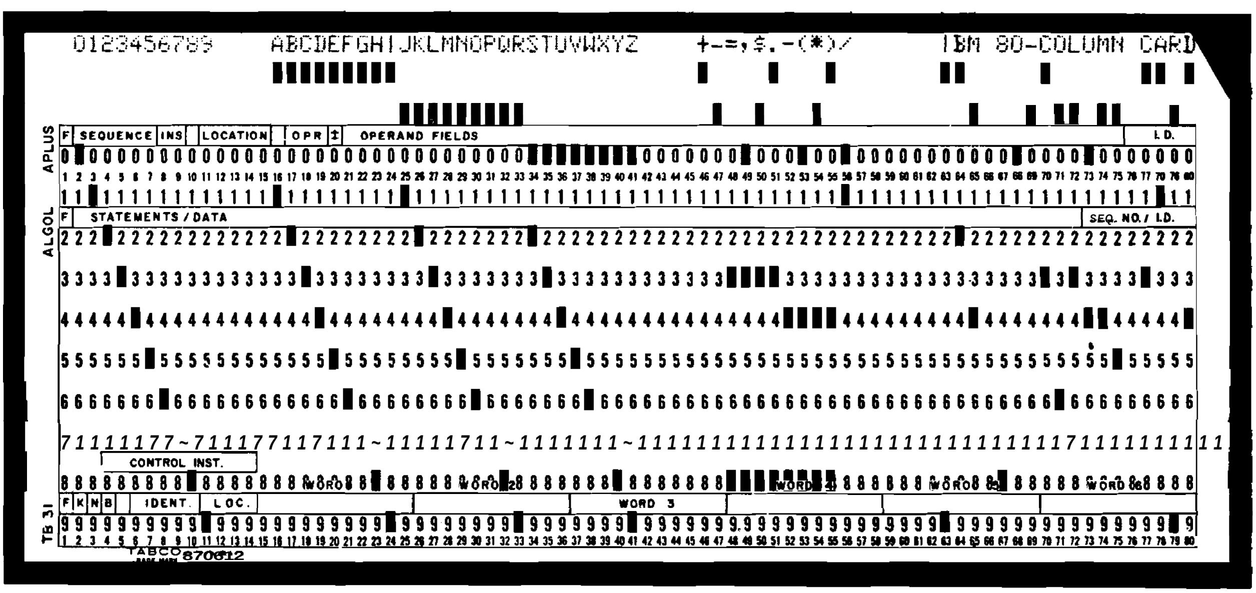


FIG. 4-2 IBM 80-column card.

forated 80-column card sent out as a bill and the 21-column stub returned by the customer with the payment. The stub (previously punched with the customer's account number and amount due) could serve as the input card for the machine to post a credit to the customer's account. Equipment modifications are available to permit the handling of any of the card sizes, but most operating systems reproduce the information from the smaller card into the largest card size as soon as the smaller card is received.[2]

Symbols are usually entered on the card by one or more of these four methods:

1. Manual keypunching
2. Automatic punching
3. Mark-sense cards
4. Pre-scored cards

In most situations, the cards are manually prepared by an operator who reads a source document and punches the data into a card, using a key punch machine (see Fig. 6-1). That is, the information is transcribed from a source document into a card. Less expensive card punches with reduced capability are available. In some cases, cards are automatically punched by computer output devices or tabulating equipment as an output, or temporary storage of the result of some machine operations.

Mark-sense cards and their associated equipment provide a capability for interpreting information that has been marked on the cards with a special pencil. Equipment is available to read these penciled marks and punch the data into the same card or into new cards. Mark-sense cards can be used in some cases to reduce the amount of manual keypunching required, but the system has found widest use where some data must be entered at the point of origin (e.g., recording physical inventories, survey results, and utility meter readings). Each marking position on the card is 3 columns wide, thus restricting the card to a maximum of 27 mark-sense characters, which may be a disadvantage in some applications. In general, any number of marking positions from 1 to 27 can be placed on a card, and in any desired position on the card. Many people with extensive data processing experience tend to avoid use of the mark-sensing technique except in those instances where there are several outstanding advantages. In general, users have indicated satisfaction with the results only under the following circumstances: there were a small number of digits to be entered; there was no pressure or requirement for speed at the time of entry; there were good physical working conditions; and the entries were to be made by clerical personnel. Most decisions to use mark-sense cards are based upon possible savings in time or cost. A cost advantage may be realized because manual keypunching and verifying is eliminated, and a time saving may be realized because the cards can be processed immediately upon their receipt, without waiting for keypunching and verifying. There does not appear to be any significant difference in the accuracies achieved by recording with mark sensing rather than by data recording and subsequent keypunching.

Pre-scored cards such as Port-a-Punch (see Fig. 4-3) permit data to be punched into the card by pushing out the perforated holes with a pencil or other probe. The pre-scored holes are positioned with a 1-column spacing between each pre-scored column. Consequently, only 40 characters, each separated by a 1-column width, can be punched into a card. The usual practice is to reproduce the information from the pre-scored cards into a standard card as soon as the card is received. This con-

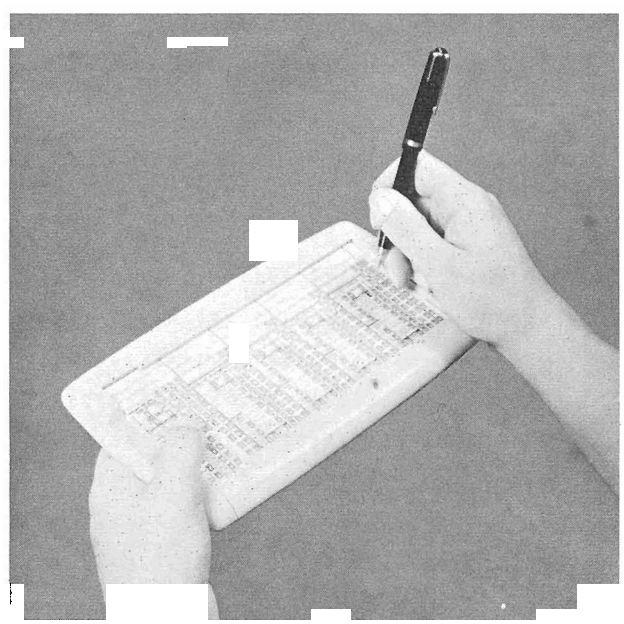


FIG. 4-3 IBM Port-A-Punch.

[2] Some very small punched tags are available from the Dennison Company for use as retail sales tags, and equipment is available to read these tags and punch the data into standard punched cards.

denses the information on the card, and removes the possibility that additional holes may pop out during subsequent machine handling. These pre-scored cards may also be useful for small, peek-a-boo file systems (see Chapt. 5).

Card sets are also available that permit duplicate papers (e.g., receipts) and punched cards to be prepared at the same time, usually at the information source. A large number of retail charge systems, such as gasoline credit cards, use this type of card set.

The Coding

Standardization* of *the Hole Patterns. As mentioned earlier, the pattern of holes in the cards is meaningless unless some definitions or standards are established by the users. There are 960 hole positions on the IBM card and 540 on the Remington Rand card. With this number of hole positions, a tremendous number of unique hole patterns are possible. However, the equipment manufacturers have adopted their own coding definitions, which utilize only a small percentage of the code possibilities. Most of the equipment recognizes only a small number of hole patterns to be valid, and usually treats any other hole patterns as errors. In some applications, such as the sorting of cards with superimposed coding on the Statistical Sorter (described later), unconventional hole patterns can be used. The restricted number of valid codes and characters (the machine language) designed into the early equipment were generally adequate for the early applications of this equipment. However, the steadily increasing application of tab card and computer equipment to a great variety of information processing tasks has created a need for more flexible coding and a greater number of valid code combinations which can be machine-manipulated. For example, sophisticated automatic programing systems for computers can be used most efficiently when the tab card codes directly represent the symbols of the programing language (e.g., $\geqq$, $\sqrt{}$, $\neq$, { }, [ ]). Some code modifications have been proposed for this purpose.[8]

[8] Voorhees, E. A., "Some Thoughts on Reconciling Various Character Set Proposals," *Communications of the Association for Computing Machinery,* Vol. 3, No. 7, pp. 408–409 (July 1960).

Bemer, R. W., "A Proposal for a Generalized Card Code for 256 Characters," *Communications of the Association for Computing Machinery,* Vol. 2, No. 9, pp. 19–23 (September 1959).

In using a set of valid characters established for a card code, there are several methods most often employed for representing information on the card:

- Direct coding
- Abbreviated coding
- Category coding
- Category coding with unconventional punching
- Superimposed coding

Direct Coding. This is the most common method and the technique that most of the equipment has been designed to utilize. With this method, there is a one-for-one correspondence between the original source data and the characters in the card columns. For example, a card used in a system for automobile registrations may contain the complete spelling of the owner's name and address as well as the vehicle information. Punching the information in this manner will allow the cards to be sorted (by name, address, or license number), or duplicated for use in separate reference files (e.g., by name, address, or license number), or listed by a printer to obtain catalogs, directories, or mailings. Telephone directories, parts catalogs, and name lists are often generated and maintained with cards that use this method of coding.

Abbreviated Coding,. In order to make efficient use of the available card columns it may be necessary to abbreviate or condense the spelling of a name, part description, subject term, or other information to be punched. Any of the methods of assigned and derived codes described in Chapt. 3 could be used to obtain these abbreviations. Cards, for example, which contain the information for addressing magazine labels or advertising material often make use of abbreviation techniques.

Category Coding. This technique is commonly employed with tab cards and edge-punched cards to make efficient use of the available coding positions. Category coding requires the use of a reference table to describe which items belong in each category so that the original terms can be correctly assigned to the proper category. The following examples illustrate how various topics might be represented as numbered categories.

Subject Field	*Code*
Physics	01
Chemistry	02
Zoology	99

Geographic Area	*Code*
San Francisco	01
Los Angeles	02
San Diego	99

Uniterm	*Code*
Aerodynamic Heating	1334
Diffraction	2123
Methyl Groups	3024

Other methods of numeric coding were described earlier, in Chapt. 3. A description of various ways to indicate relationships between descriptors coded in different areas of a card has been given by Dekker.[4]

Category Coding with Unconventional Punching. Most punched card equipment recognizes only a small group of punch patterns as being valid codes, and rejects all other patterns as illegal. The processing equipment either interprets these invalid patterns as errors or manipulates the information to provide an incorrect result. Printing or tabulating equipment does not have any characters to correspond to the invalid punches, and key punch machines cannot conveniently be used to punch the invalid patterns into the card. However, some of the sorting equipment can operate with the invalid punches.

Because the use of non-standard punch patterns (e.g., a 7 and an 8 punched in the same column) will provide a greater number of coding possibilities for each card column, the use of the invalid patterns provides a more efficient coding scheme than could be obtained with direct or category coding with conventional punching. In a strict sense, it is possible to uniquely represent 2^{12} or 4096 different topics in one IBM card column by using this technique. However, for convenience the user may wish to employ only those code combinations which utilize 2 or 3 punches per column. This will decrease the number of total coding possibilities, but this number may still be much larger than that obtained with conventional coding. Unconventional punching might be used, for example, to permit a single column to represent many different animals, such as:

Animal	Rows Punched in a Column
Horses	3 and 4
Cows	3 and 5
Sheep	1 and 6
Cats	7 and 8

The codes might also be assigned in a systematic manner to represent a hierarchical structure in a single column. For example:

Subject	Rows Punched in a Column
Airplanes, all types	1
Airplanes, jet	1 and 2
Airplanes, turboprop	1 and 3
Airplanes, propeller	1 and 4
Airplanes, helicopters	1 and 5
Ground vehicles, all types	2
Ground vehicles, 4 axles	2 and 3
Ground vehicles, 6 axles	2 and 4
Ground vehicles, 8 axles	2 and 5
Ground vehicles, more than 8 axles	2 and 6
Ground vehicles, tracked	2 and 7
Boats, all types	3
etc.	etc.

Superimposed Coding. It is possible to superimpose descriptions or indexing information into fixed fields of a tab card.[5] This process will usually generate some invalid punch combinations and introduce the inconveniences of invalid punch combinations men-

[4] Dekker, J., "Interfiling of Descriptors for Row-by-Row Coded Punched Card Machine Literature Searching Systems," Chapt. 8 in ***Information Retrieval and Machine Translation,*** Part I, A. Kent, editor (Interscience Publishers, New York, 1960).

[5] Anon., "Superimposable Punched Cards as a Means of Reference to Periodicals," ***Unesco Bulletin for Libraries,*** Vol. 12, No. 10, pp. 228–229 (October 1958).

Weil, B. H., and E. A. Clapp, "The Adaptation of the ASM-SLA Metallurgical Literature Classification for Use in a Machine-Sorted Punched Card System," Chapt. 20 in ***Information Systems in Documentation,*** J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

tioned earlier. The use of the superimposed punching will normally restrict the card-handling operations to the use of the IBM Statistical Sorter (described later, in Chapt. 6).

PAPER TAPE

Punched paper tape is available in many different sizes and is used with many different code patterns. The code may be represented by holes that are completely punched out (chad tape) or partially punched out (chadless tape) or printed but not punched (e.g., Burroughs Electrographic tape). The holes may be sensed with mechanical feelers (for chad and chadless tape), sensed photoelectrically by light beams transmitted through the holes (chad tape), or sensed by beams reflected from the printed spot (Electrographic tape). Tape that is sensed by photoelectric transmission usually requires the use of an opaque paper to reduce the amount of light transmitted directly through the paper stock, but there are no other restrictions on the color of the paper. For infrequent usage and the slow-speed handling of most applications, a relatively light and inexpensive paper stock can be used. However, if the paper tape is to receive any high-speed machine handling (e.g., 100 inches per second), or repeated use as a permanent tape in a control system, then it should be made from suitably durable stock. In special instances, plastic tape (such as Mylar tape) has been used for this purpose. Paper tape is usually used in a continuous reel form; however, unit-record card stock and continuous fan-fold tape are also used. The edge of a paper document (e.g., an 8%-by-11-inch page or a 3-by-5-inch card) could be punched in such a way as to permit the body of the paper to be used as a document, while the edge punching was used to carry some information for machine handling.

The basic families of paper tape codes are referred to as 5-, 6-, 7-, and 8-channel codes.[6] The "channel" refers to a coding line running along the length of a tape. A 5-channel tape, for example, has a pattern of 5 hole positions across the width of the tape. The hole combinations across the width of the tape determine the variety of codes possible. One row of hole positions across the width of a tape is called a "frame" and represents a single character, and is analogous to an IBM card column. Several codes are illustrated in Fig. 4-4. As with

[6] Tholstrup, H. L., "Perforated Storage Media," Electrical Manufacturing, Vol. 62, No. 6, pp. 53–61, 276 (December 1958).

Tholstrup, H. L., "Numerical Control by Punched Tape," Instruments and Control Systems, Vol. 34, No. 9, pp. 1643–1646 (September 1961).

Anon., "All About Paper Tape," Datamation, Vol. 5, No. 3, pp. 8–12 (May–June 1959); Vol. 5, No. 4, pp. 48-51 (July–August 1959).

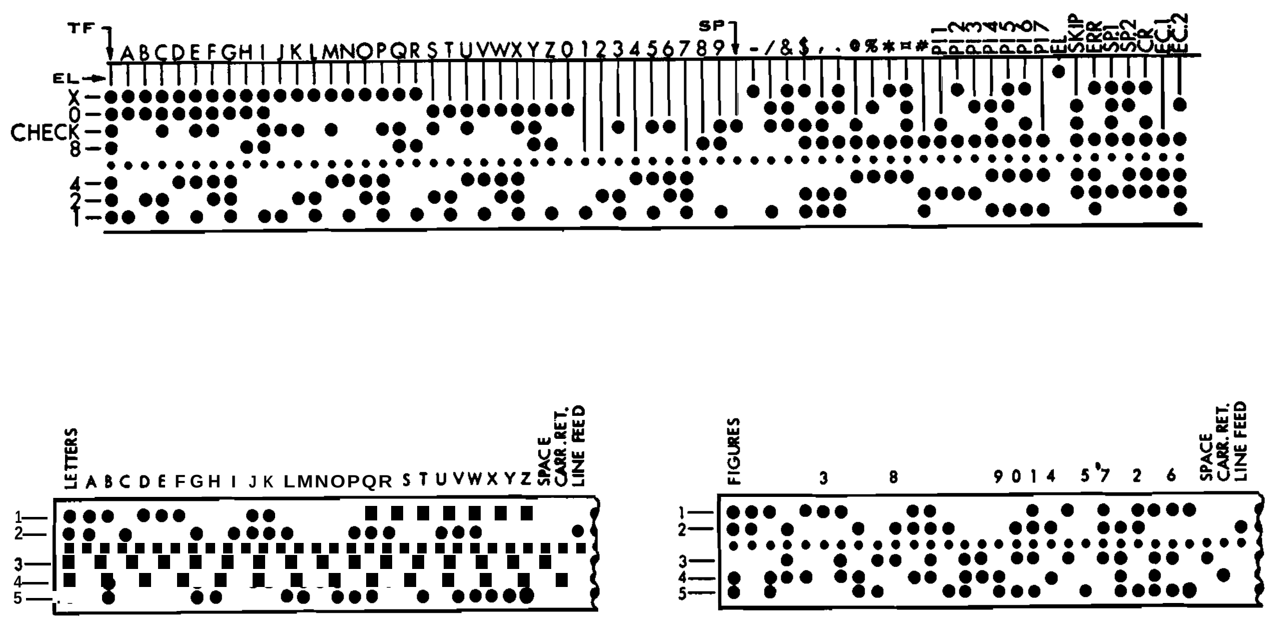


FIG. 4-4 Representative paper tape codes.

the tab cards, no single standard code exists, although several of the codes (notably the teletype and IBM codes) are used much more freauentlv than the others. The purchaser of the paper tape equipment usually has the final choice of codes. To handle code compatability problems, equipment is available to read a tape that is punched with one code and use this information to punch a tape in another code.? None of the common paper tape typewriter systems print symbols on the paper tape to correspond to the punched information, although the practice is followed by some teletype equipment.

MAGNETIC TAPE

Information is stored on digital magnetic tape and magnetic cards by forming discrete-variable magnetic flux patterns in magnetic media according to some pre-determined code. The patterns are formed by passing an electric current through a recording head which has been placed very close to the surface of the magnetic medium. As the electric current passes through the recording head in a prescribed sequence, a corresponding sequence of magnetic field changes will occur. The magnetic material is held close enough to the recording head so that the magnetic material will be magnetized in the same pattern as the magnetic field of the original electric current. During recording, the magnetic medium is moved rapidly past the recording head so that new information is not written over old information. The materials used for the storage medium are such that they will permanently retain the flux patterns until the medium is remagnetized by another recording operation, or by some other magnetic field. As with any magnetic device, the medium will lose its magnetic patterns. and hence its memory, if it is subsequently exposed to high temperatures, extreme physical shocks, or large magnetic fields.

The magnetic pattern is normally invisible to the eye, so that recordings usually must be sensed electrically instead of visually. The sensing of the tapes is done with a reading head, which is similar to the writing head. When the storage medium is moved past the read head, the stored flux patterns of the medium induce a voltage in the head, which tends to follow the pattern of the stored flux. In this way, an electrical signal is obtained that corresponds to the recorded flux pattern. At this point, a clarification of the term "digital recording" would seem to be in order. Digital recording means that only discrete numbers or digits are recorded, most often with a binary number system. The information is represented by patterns of discrete-valued variables. The flux patterns for each coding position are usually oriented only in one of two allowed states, and numeric and alphabetic symbols are represented by uniquely specified groups of flux patterns. The "track" of the magnetic tape is analogous to the "channel" of the paper tape, and "frame" has the same meaning for both types of tape. Figure 4-5 shows a section of magnetic tape with 7 tracks or channels. Many magnetic tape systems have a 5-, 6-, or 7-channel code similar to that used with punched paper tape. In contrast, analog recording allows a continuous variation in the flux pattern in a manner that is analogous to the variations in the original stimulus. An audio tape recorder, for example, provides a recording in which the flux pattern is analogous to the original sound pattern. With a few noted exceptions, magnetic recording will, in this book, imply digital recording.

? Converter equipment is also available to read paper tapes and to transcribe the information onto other storage mediums, such as punched cards or magnetic tape. Conversely, equipment is available to transcribe information from many different mediums onto punched paper tape.

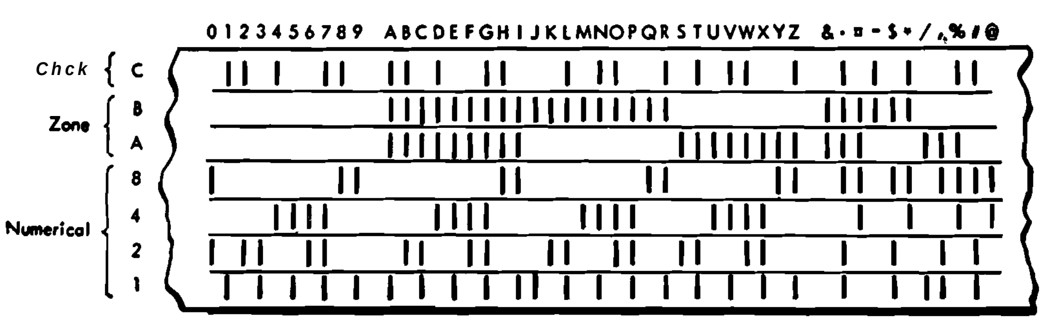


FIG. 4-5 *Representative computer magnetic tape coding.*

A roll of magnetic tape is the usual storage medium, although magnetic cards have been used.[8] The magnetic tape normally consists of rolls of thin plastic (such as Mylar) ribbon that has been coated with a very thin layer of magnetic oxide material. In some cases, a plastic lamination is placed on top of the oxide (a "sandwich tape") to prevent the oxide from flaking or rubbing off during reading or writing. The rolls are normally 2400 to 3600 feet in length, and ½ inch, ¾ inch, or 1 inch in width. It is usually necessary to store and use the tape in a dust-free and controlled environment (nominally 50 to 85°F and 50 to 90% relative humidity) to prevent damage to the tapes or degradation of the information on the tape. If the tape is too dry, it tends to build up a charge of static electricity that may cause it to physically jam the tape transport mechanism. Combinations of moisture and high temperatures will cause the tape to stretch out of shape when it is being pulled at high speeds.

Since the information can be recorded in very small regions on the tape, it is possible to record several parallel tracks on the same tape. Figure 4-5 shows the coding and recording pattern of a representative digital magnetic tape. The information or pulse patterns can be stored very compactly on tape; typically there are from 5 to 20 tracks per tape, and from 100 to 1000 pulses per linear inch of track. The tape can also be moved at very high rates of speed, 75 to 200 inches per second. The high recording density and tape speed result in a rapid information transfer rate of 25 to 200 thousand digits or characters per second. And as with the punched card and paper tape systems, the physical dimensions, recording method, and coding system are parameters whose values have been established separately by almost every equipment manufacturer. Consequently, no single standard tape or coding system exists for digital magnetic tape. Only in a few instances can the magnetic tapes written on one manufacturer's computer be read on another manufacturer's computer.

MACHINE REPRESENTATION OF NATURAL LANGUAGES

Text Material

It would be most useful to have machines that could read ordinary English text, such as newspapers, books, and journal articles. Complex processing of text material—such as automatic indexing, automatic abstracting, machine translation, concordance construction, and automatic editing—cannot be done effectively or economically unless the processing equipment can store, recognize, and process the many different symbols and forms as used and recorded in our natural language. Present-day equipment is generally inadequate in this respect. Whereas the text of a published dictionary may currently use an alphabet of approximately 1400 different characters (uppercase and lowercase letters of 4 different type styles plus special characters), most data processing equipment is only designed for a total alphabet size of something less than 90 characters. For this reason, many coding tricks and simplifications must be used to represent text material in machine language today. Several schemes have already been proposed to represent text material in the language of the current data processing machinery.[9]

Image Material

Chemical structures (such as molecules and polymers) and their symbolic notation or coding are a part of our natural language that is usually difficult to represent in machine language. Several coding procedures have been developed: [10] semi-automatic

[8] The magnetic cards will be discussed later (Chapt. 8) in conjunction with the Magnacard system.

[9] Nugent, W. R., "A Machine Language for Document Transliteration," paper presented at the 14th National Conference of the Association for Computing Machinery, Massachusetts Institute of Technology (September 1959).

Newman, S. M., R. W. Swanson, and K. C. Knowlton, "A Notation System for Transliterating Technical and Scientific Texts for Use in Data Processing Systems," Chapt. 7 in *Information Retrieval and Machine Translation, Part 1* (Interscience Publishers, New York, 1960).

[10] Waldo, W. H., and M. DeBacker, "Printing Chemical Structures Electronically: Encoded Compounds Searched Generically with IBM-702," *Proceedings of the International Conference on Scientific Information*, Vol. 1, pp. 711–730 (National Academy of Sciences, Washington, D.C., 1959).

Opler, A., and N. Baird, "Display of Chemical Structure Formulas as Digital Computer Output," *American Documentation*, Vol. 10, No. 1, pp. 59–63 (January 1959).

Opler, A., and N. Baird, "On the Automatic Manipulation of Repreeentations of Chemical Structures," *American Documentation*, Vol. 10, No. 2, pp. 130–134 (April 1959).

Opler, A., "A Topological Application of Computing Machines," *Proceedings of the 1966 Western Joint Computer Conference*, pp. 86–88.

Wheeler, K. W., "A Structure Code for Organic Compounds," *American Documentation*, Vol. 9, No. 3, pp. 198–207 (July 1958).

procedures have been developed for translating inorganic compound names into formulas,[11] for translating structural formulas of organic compounds into a linear notation,[12] and for converting compound names into molecular formulas.[13]

Graphic representations (e.g., pictures, maps, drawings, equations) are also a part of our natural language, and will eventually be handled and processed with machines—just as text material is processed today. Graphic images may have to be stored for file-searching applications involving such things as: (1) engineering drawings, (2) curves, figures, and equations associated with document abstracts, (3) copies of legal documents and signatures associated with business records, (4) property descriptions to accompany tax records or title files, (5) chemical structures to accompany their associated compound characteristics. The method most commonly employed for the digital storage of an image is to approximate the image with a rectangular matrix of discrete, equal-area elements, each element having a quantized intensity or shading corresponding to that of the original image. The accuracy of the approximation depends upon such things as the number and size of matrix elements used and the amount of tone scale preserved.

For some situations it is possible to directly and uniquely describe an image by a single linear notation or number derived from an examination of its topology or general configuration. However, the application of this type of coding is necessarily restricted to a very special class of well-structured images, such as chemical structures, electrical circuit diagrams, and flow charts or block diagrams.[14] An automated program has been developed to derive an identifying statement or coded description of any simple pattern.[15]

The many different image processing operations described to date can be grouped into a few general categories:

Derivation of image statistical data
Blob or area counting
Center-of-gravity computation
Translation, rotation, compression, and expansion
Superimposition and generation of images
Image smoothing, thickening, and thinning
Other contour enhancement techniques
Topological determinations and operation3

The input and output of graphic data (images) have been handled for digital computers in a variety of ways, of which flying spot scanners and cathode ray tube displays are the most sophisticated. The processing has been programed or simulated on several different computer types such as the National Bureau of Standards SEAC; MIT's Whirlwind I, Memory Test Computer, TX-O and TX-2; and the IBM 650,704,709, and 7090 systems. A large number of processing techniques have been developed, so that it should be possible to construct a fairly comprehensive library of processing routines for application of computer systems to general image processing problems and research. However, there have been few instances, reported in the literature, of practical application of these techniques to real operational problems. Image processing techniques could be used effectively in a number of applications: (1) character and pattern recognition, and

[11] Sayfer, A. L., and V. S. Shteyn, "An Algorithm for Converting the Name of a Complex Compound Presented in Rational Nomenclature into a Linear Formula," in ***Foreign Developments in Machine Translation and Information Processing,*** No. 58, JPRS: 11483 (December 12, 1961). Distributed by Office of Technical services, US. Dept. of Commerce, Washington, D.C. This is an English Translation of a 1960 Soviet article.

[12] Borshchev, V. B. et al., "An Algorithm for Translating Structural Formulas in Organic Chemistry into Canonical Notation," in ***Foreign Developments in Machine Translation and Information Processing,*** No. *58,* JPRS:11483 (December 12, 1961). Distributed by Office of Technical Services, U.S. Dept. of Commerce, Washington, D.C. This is an English Translation of a 1960 Soviet article.

[13] Garfield, E., ***An Algorithm for Translating Chemical Names to Molecular Formulas,*** a report of the Institute for Scientific Information, Philadelphia, Pennsylvania (1961).

[14] Luhn, H. P., ***Identification of Geometric Patterns by Topological Description of Their Envelopes,*** Technical Report No. 001.011.599, IBM Corp., Product Development Lab., Poughkeepsie, New York (April 1956).

Opler, A., and N. Baird, "Display of Chemical Structure Formulas as Digital Computer Output," ***American Documentation,*** Vol. 10, No. 1, pp. 59–63 (January 1959).

Opler, A., and N. Baird, "On the Automatic Manipulation of Representations of Chemical Structures," ***American Documentation,*** Vol. 10, No. 2, pp. 130–134 (April 1959).

Opler, A., "A Topological Application of Computing Machines," ***Proceedings of the 1966 Western Joint Computer Conference,*** pp. 86–88.

Wheeler, K. W., "A Structure Code for Organic Compounds," ***American Documentation,*** Vol. 9, No. 3, pp. 198–207 (July 1958).

[15] Grimsdale, R. L., et al., "A System for the Automatic Recognition of Patterns," ***Proceedings of the Institution of Electrical Engineers*** (London), Vol. *106,* Part B, No. 26, pp. 210–221 (March 1959).

Grimsdale, R. L., "Automatic Pattern Recognition," ***Wireless World,*** November 1959, pp. 499–501.

simulation of character and pattern recognition logic; (2) automatic encoding of chemical structures, electrical circuits, or other line drawings; (3) object counting, such as particle or bacterial colony counts; (4) determination of areas of selected terrain or vegetation on aerial photographs; (5) analysis of tracking data such as photos of satellite tracks against a star background; (6) analysis of astronomy photographs to compute positions and motions, and to establish star catalogs; (7) generation of elevation contour lines from aerial photographs; (8) selection of optimum road routing and cut-and-fill computations from contour maps; (9) superimposition of images to form composite images; (10) improvement of contrast on such low-definition images as X-ray photographs, and the enhancement of these images to accentuate outlines and skeletal features. This is already being done for the pictures obtained by some U.S. space satellites. It seems reasonable to expect that computer equipment will approach most of these other capabilities with some success in the next 5 or 10 years.

Additional References

Bemer, R. W., "Survey of Coded Character Representation," *Communications of the Association for Computing Machinery,* Vol. 3, No. 12, pp. 639–642 (December 1960).

Bonnett, H. T., and D. W. Calhoun, "Application of Line Formula Notation in an Index of Chemical Structures," *Journal of Chemical Documentation,* Vol. 2, No. 1, pp. 2-6 (January 1962).

Doyle, W., "Recognition of Sloppy, Hand-Printed Characters," *Proceedings of the 1960 Western Joint Computer Conference,* pp. 133–142 (Institute of Radio Engineers, New York).

Frishkopf, L. S., and L. D. Harmon, "Machine Reading of Cursive Script," Chapt. 23 in *Information Theory,* C. Cherry, editor (Butterworth, London, 1961).

Garfield, E., "Information Theory and Other Quantitative Factors in Code Design for Document Card Systems," *Journal of Chemical Documentation,* Vol. 1, No. 1, pp. 70–75 (January 1961).

Gull, C. D., "Holes, Punches, Notches, Slots, and Logic," Chapt. 19 in *Punched Cards,* 2nd ed., R. S. Caaey et al., editors (Reinhold Publishing Corp., New York, 1958).

Hayward, H. W., *A New Sequential Enumeration and Line Formula Notation System for Organic Compounds,* U.S. Patent Office, Washington, D.C., Research and Development Report 21 (November 1961).

Kirsch, R. A., et al., "Experiments in Processing Pictorial Information With a Digital Computer," *Proceedings 1967 Eastern Joint Computer Conference,* pp. 221–229 (Institute of Radio Engineers, New York).

Neisser, U., and P. Weene, "A Note on Human Recognition of Hand-Printed Characters," *Information and Control,* Vol. 3, No. 2, pp. 191–196 (June 1960). See also AD-236 395.

Oliver, B. M., "Efficient Coding," *Bell System Technical Journal,* Vol. 31, No. 4, pp. 724–750 (July 1952).

Rosen, L., "Characteristics of Digital Codes," *Control Engineering,* December 1959, pp. 115–119.

Ross, H. M., "Considerations in Choosing a Character Code for Computers and Punched Tapes," *Computer Journal* (British), Vol. 3, No. 4, pp. 202–210 (January 1961).

Ross, H. M., "Further Survey of Punched Card Codes," *Communication of the Association for Computing Machinery,* Vol. 4, No. 4, pp. 182–183 (April 1961).

Smith, H. J., and F. A. Williams, "Survey of Punched Card Codes," *Communications of the Association for Computing Machinery,* Vol. 3, No. 12, p. 638 (December 1960).

Starker, L. N., and J. A. Cordero, "The Cyanamid Organic Structure Code and Search System," *Journal of Chemical Documentation,* Vol. 2, No. 1, pp. 12–15 (January 1962).

Waldo, W. H., "Searching Two-Dimensional Structures by Computer," *Journal of Chemical Documentation,* Vol. 2, No. 1, pp. 1–2 (January 1962).

Edge-Notched Cards
Interior-Notched Cards

– 5

Manual card systems

Manual card systems such as the 3-by-5-inch catalog card files have been in use for many years, and have proven to be a very flexible means for recording, sorting, merging, re-arranging, and updating files in which each file item or unit-record consists of a relatively small amount of information and can be handled independently of all the other file items.[1] Some mechanical aids (e.g., loose leaf sorters or sorting bins) have been developed to assist in the manipulation and re-arrangement of the cards, but all of the major operations are still performed manually.

In order to speed up or simplify the manual handling of these cards, some mechanical modifications have been incorporated; namely some special means of coding, notching, and selecting the edges or interiors of the cards. This chapter discusses two general types of manual card systems that have been modified in this manner to reduce effort and make the unit-record more flexible: the edge-notched card and the interior-notched card. A description of the coding methods and equipment employed, and suggestions for how these cards might be used to advantage, are included. Both the edge-notched and interior-notched systems are basically manual systems, although some simple mechanical aids are available to augment both types of systems.

[1] Weeks, B. M., *How to File and Index,* rev. ed. (Ronald Press, New York, 1951).
Odell, M. K., and E. P. Strong, *Records Management and Filing Operations* (McGraw-Hill Book Co., New York, 1947).

EDGE-NOTCHED CARDS

This section is concerned with card forms which are coded with notches along the edges of the card (see Fig. 5-1). The notches are usually imposed on pre-punched pilot holes or guide lines. These cards are commonly referred to as "marginal punched cards," "edge-punched cards," "edge-notched cards," "manual punched cards," and "notched cards." There are several types and forms of cards, furnished by several different equipment and forms manufacturers. The card system most commonly employed is the Keysort Card, marketed by the Royal McBee Corp. Table 5-1 lists many of the types of edge-punched cards that are commercially available. Several representative notched cards are shown in this chapter to illustrate specific points brought up in the discussion.

The notched card is especially useful for two types of jobs: (1) *sequence sorting to* arrange data or records into a logical sequence to simplify tabulation of like items or the use of the items in a predetermined sequence; and (2) *selective sorting* to select those records which fall into a single specific category (direct selection), or to select those records which fall into a specified combination of categories (multiple selection).

The Storage Medium

The individual notched card provides a very flexible storage and retrieval system, which can be tailored to fit the requirements of each application.

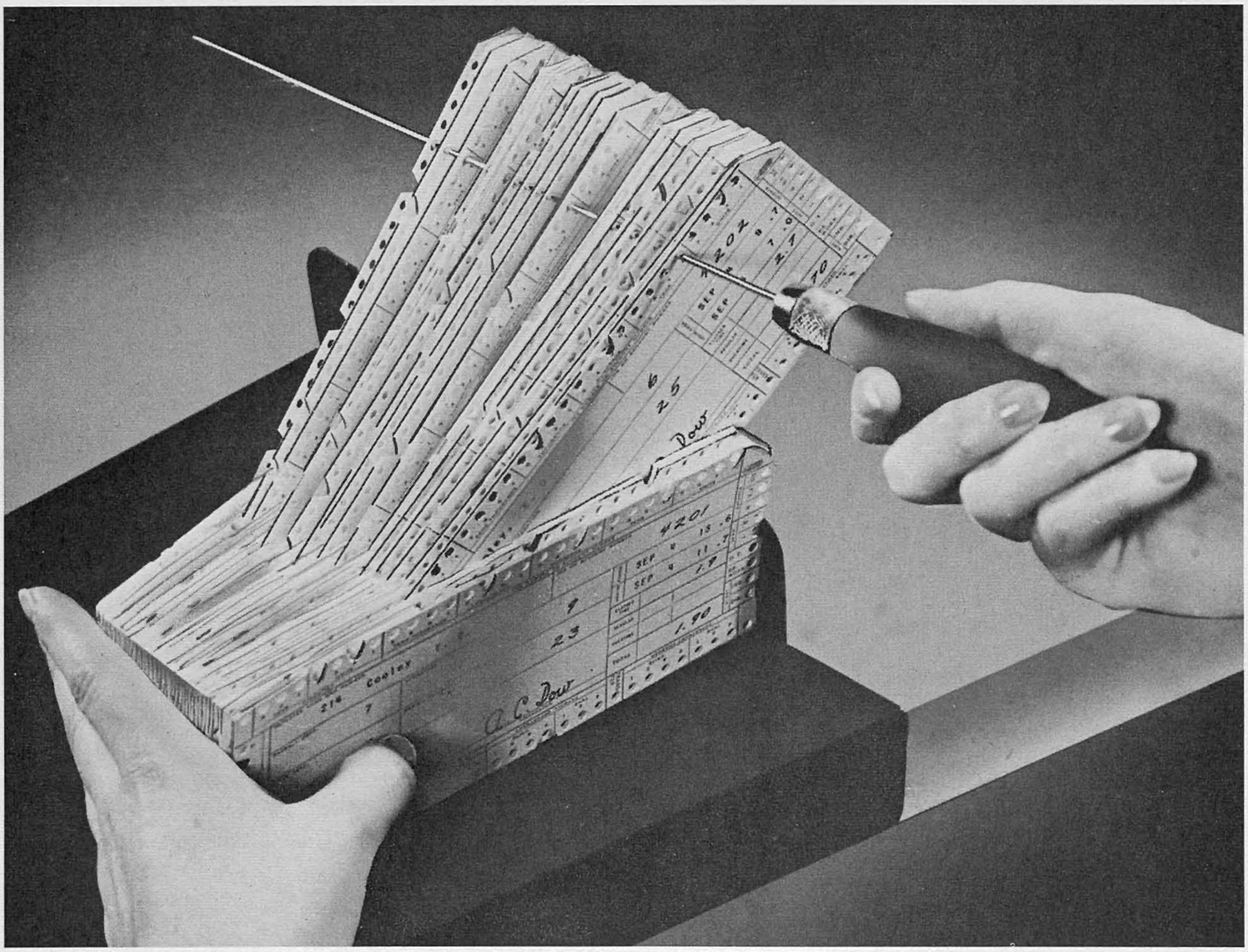


FIG. 5-1 Card selection with an edge-notched card system.

The exact shape of the card and the number of coding positions on the edge of the card are usually determined by each card supplier. The coding system used on the card is usually developed by each user and pre-printed on the cards, although some standard preprinted cards are available from each forms supplier.

Regardless of the size of the card or the type of coding used, generation and search are similar in almost all systems. Portions of the card edge are notched so that the code position is no longer surrounded by card stock and the card cannot be supported by a needle which is passed through that code position. The cards may be notched with a simple hand punch (see Fig. 5-2), in which case it is useful to employ a coding template.[2] Other mechanical equipment is also available for card notching (see Figs. 5-11 through 5-15). The searching is performed for most notched card systems (see Fig. 5-1) by passing a needle through the appropriate pilot hole position in a bundle of cards and allowing the notched cards to fall free from the un-notched cards.[3] Because of the friction between adjacent cards, the cards must be separated, fanned, or vibrated during the search operation, to ensure that no notched cards are kept from falling because of the friction forces. One modification of the use of pilot holes is employed with the Zatocard (see Fig. 3-9), which uses printed marks but no prepunched holes on the edges of the cards. The Zatocard coding is achieved by notching the cards in the indicated positions, and the searching is done by

[2] Gould, D. W., "The Use of Templates with Punched Cards in the Code Indexing of Technical Literature," *Special Libraries,* Vol. 50, No. 9, pp. 451–453 (November 1959).

[3] Casey, R. S., and J. W. Perry, "Elementary Manipulations of Hand-Sorted Punched Cards," Chapt. 2 in Punched Cards, 2nd ed., Casey et al., editors, cited previously.

TABLE 5-1

Commercially Available Notched Card Systems

Card Name	Distributor
E-Z Sort	E-Z Sort Systems, Ltd. 45 Second Street San Francisco 5, California
Keysort	Royal McBee Corp. Port Chester, New York
Zatocard	Zator Co. 140½ Mt. Auburn Street Cambridge 38, Massachusetts
Unisort System	Todd Co. Div., Burroughs Corp. 950 Charter Street Redwood City, California
Flexisort System	Superior Business Machines, Inc. 285 Madison Avenue New York, New York
Needlesort System	Arizona Tool and Dye Co. 31 East Rillito Street Tucson, Arizona
Practa Data-Card	Practa Data-Card System 1245 East Walnut Street Pasadena, California
Electrofile	Acme Visible Records, Inc. Crozet, Virginia

supporting the deck of cards on long rods which have been placed directly beneath the specified positions and at right angles to the card (see Fig. 5-3). In this way, cards which satisfy the search criteria will drop down, and the notches instead of the card edges will be resting on the rods. The rods support the unwanted cards on their edges instead of the interior of the notch, so that they sit higher than the others. Consequently the unwanted cards can be removed by skewering them through a hole provided in the top of the cards, and lifting them away from the selected cards.

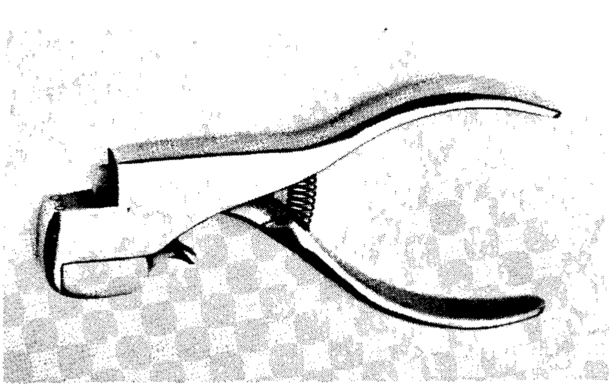


FIG. 5-2 Simple manual punch.

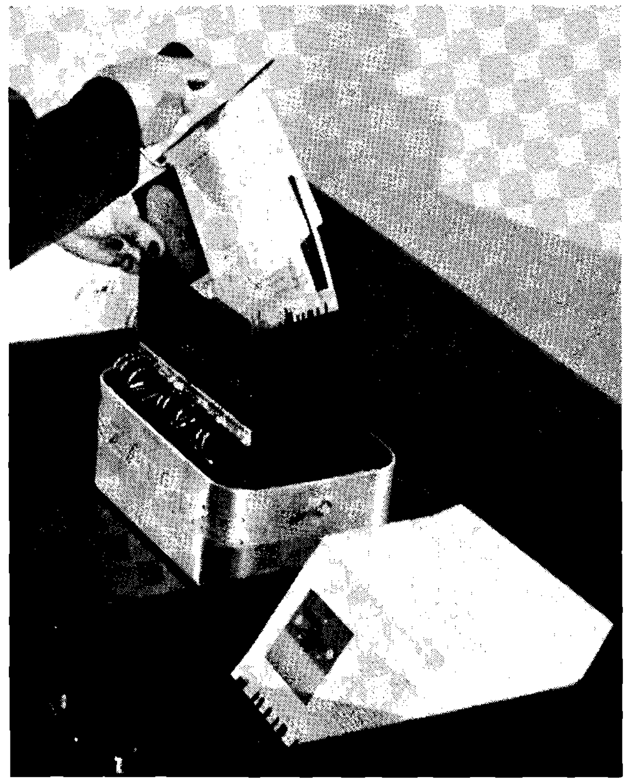


FIG. 5-3 Card selection with a Zatocard selector.

The cards are available in a large variety of sizes, shapes, colors, and number of coding positions. Special cards are also available to hold duplicating stencils, printing plates, or microfilm inserts with each card, so that printing can be done directly from the selected card. Some card stock is also available in sets with carbon paper or carbon backing so that multiple copies of the original information can be generated for filing or distribution. Card stock is available in the form of coupon books, which are useful for, say, loan repayments or toll tickets. For these applications, information common to all of the cards can be notched before the set of cards is issued. In some situations, it may be convenient to use a sorting field on just a small portion of a document. An example of this is the use of a punch position for account numbering on commercial bank checks. For the inevitable mis-punched hole or change in classification, gummed stickers (Card Savers) are available to patch the holes and restore the edge of the card. Card Savers may also be used to fasten cards together when they are to be handled as a set.

Information may be entered onto the card by handwriting, or typing, or office duplicating equipment (spirit, offset, xerographic). For some applications, it may be convenient to print some information on the notched card from the same duplicating master that was used to prepare some other form or document. This might be useful, for example, when the cards are used as part of a production control system in which common information (e.g., order number and description) is printed onto the notched cards used for labor and requisition charges as well as the other accompanying shop documents.

One very interesting modification of the edge-notched card is the Practa Data-Card (see Fig. 5-4). The Practa Card is actually a folded piece of paper which has 21 notches in the center fold, with each notch constructed in such a way that it can hold a small plastic tab or signal that protrudes from the edge of the paper when the paper is folded. These plastic tabs come in different shapes and 14 different colors to act as markers on the card. In addition, plastic tags are available which are made in the form of a loop. Using the looped tags, the deck may be needled for searching (see Fig. 5-5) in much the same way as edge-notched cards are needled. The loops are offset from the center of the notch so that a loop can be put on either side of a notch. In practice this provides 34 coding positions for the loops. The cards may be recoded by removing and re-positioning the tags. The tags do not fall out when the card is opened up for reading or insertion in a typewriter.

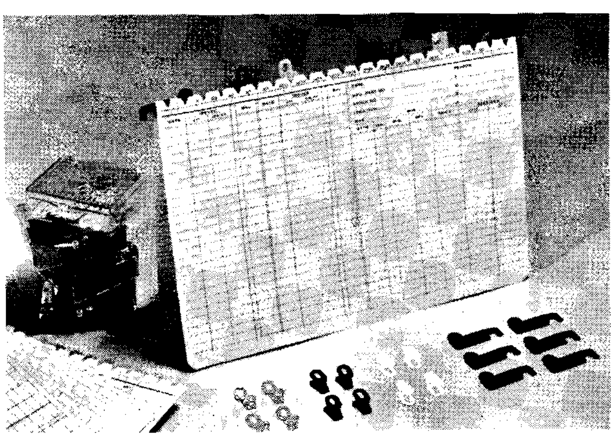


FIG. 5-4 Practa Data-Card.

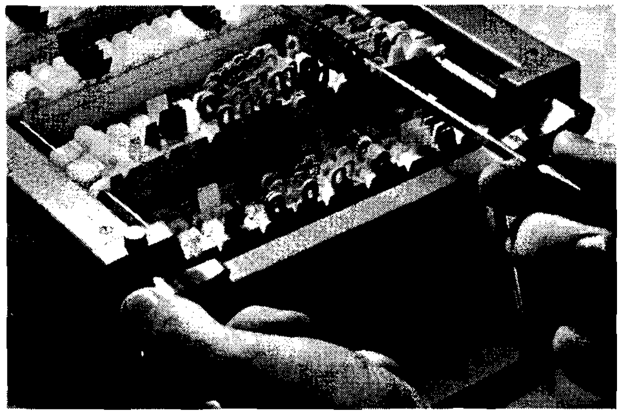


FIG. 5-5 Needling a deck of Practa Data-Cards.

The Coding

Direct Numeric *Information.* Many applications require that some data, such as an employee number or part number, be punched directly into the card. There are several coding techniques which permit direct representation of such numeric or alphabetic information. However, all of these methods for direct representations, especially the alphabetic, require a relatively large amount of coding space. Since the card has a limited number of coding positions along the edge, more conservative coding methods are usually used.

Figure 5-6 illustrates several schemes for representing numeric data by punch patterns. The schemes which assign a separate hole to each of the ten digits (Fig. 5-6a) are the easiest to use, but require the most coding space. With double-row punching (i.e., an inner and an outer row of holes) or triple-row punching, each of the holes can be uniquely assigned, with a resulting saving in card space (Fig. 5-6*b*, c). Cards with two, three, four, or more rows of punching can efficiently code large numbers by assigning each row to a specific digit position (e.g., hundreds, tens, units) and including all ten possible digits in the coding field (see Fig. 5-6*c*). With this method, some provision must be made to distinguish duplicate digits in the same number. Schemes which allow a particular hole to

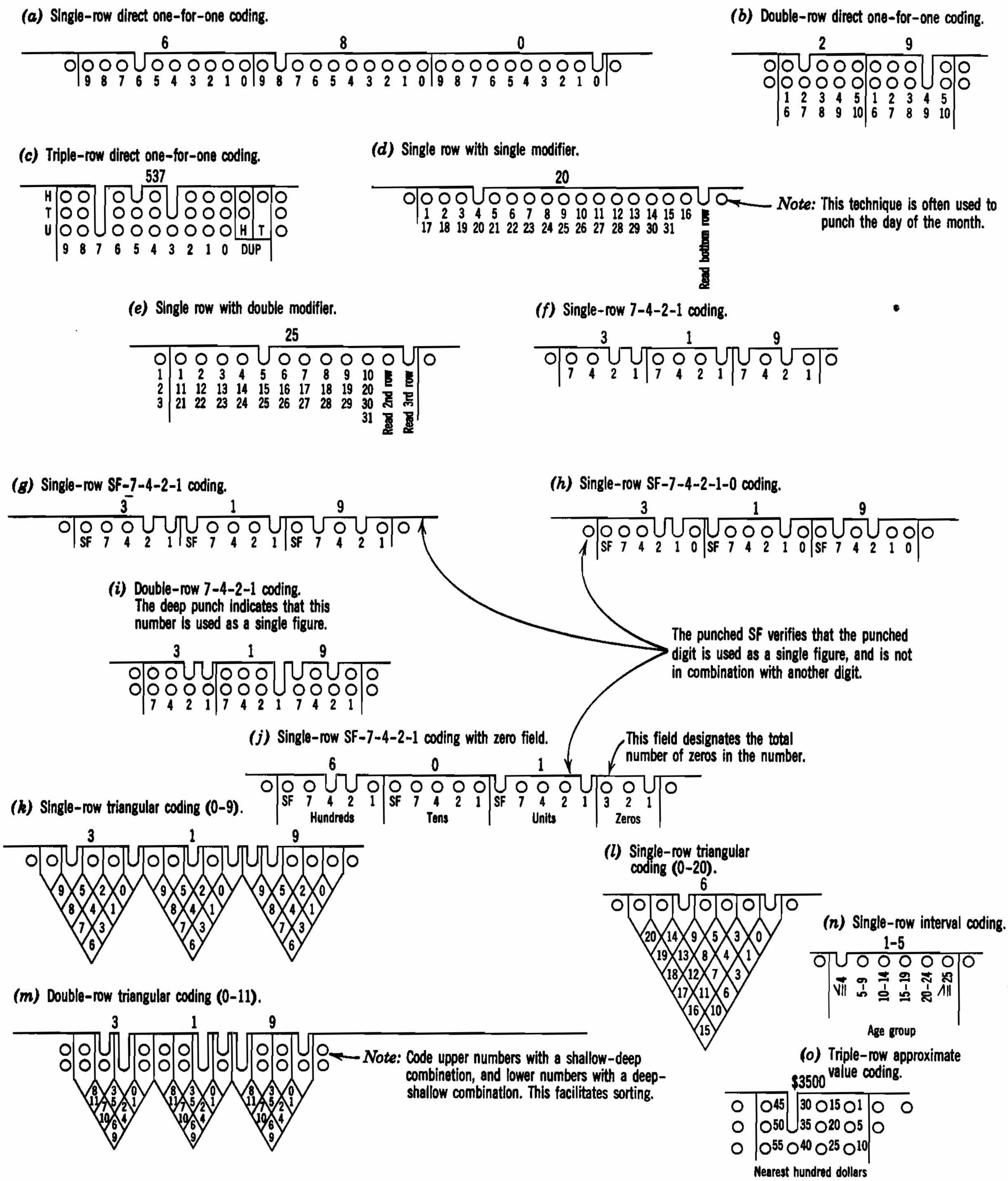


FIG. 5-6 Methods for direct coding of numeric data.

assume more than one meaning by using one or more modifying code positions to assign the meanings are somewhat more efficient and almost as easy to use (Fig. 5-6*d*, *e*).

From a space standpoint, one of the most efficient and most commonly used schemes is the 7-4-2-1 code, which only requires four holes to represent any number between 0 and 14 (Fig. 5-6*f*). Normally the four holes are only used to represent the digits 0 to 9, with zero represented by an unnotched code field. This introduces some ambiguity since a field with no punches in it may represent either missing data or a zero digit. The 1 hole will be punched every time that the numbers 1, 3, 5, or 8 are punched. If there was a need to search the file to select the cards which were punched with a 1, then a single needle pass in the 1 hole would also drop out all the cards which were punched with a 3, 5, or 8. Similarly, a single needle pass in the 4 hole would also select the cards notched with 5's and 6's. For this reason, this particular code does not lend itself to rapid selective searches. However, it is extremely useful for applications that require sorting cards into sequence. A one-digit field (0 to 9) in the 7-4-2-1 code can be completely sorted into numerical order with just four needle passes, whereas a one-digit field (0 to 9) in a card with direct one-for-one coding will require ten needle passes. This sorting into numerical sequence with the 7-4-2-1 code can be performed very simply by needling the 1, 2, 4, 7 holes in that sequence and putting the fallen cards (in the same order that they had in the deck before they fell) at the rear of the deck after each pass and before the next pass is made.

The addition of one or two extra positions to the 7-4-2-1 code reduces the ambiguity and permits the decks to be selectively searched more rapidly, while still permitting relatively easy sequential sorting (Fig. 5-6*g*, *h*). A 0 hole position is used to punch the digit 0, and an SF position is used to indicate that the punched digit position is used as a "single figure" and not in combination with another digit position. Thus the joint or sequential needling of the 1 hole and the SF hole will select only the cards which are punched with the value 1 and will ignore the other cards which use the 1 hole, such as the 3, 5, and 8. For this scheme two needle passes, in sequence or simultaneously, are required to select a card with a specific digit. This technique guarantees, however, that no extraneous cards will be selected by the dual needling operation. For sorting into numerical sequence the SF and 0 holes are ignored and the deck is sorted in the manner described for the 7-4-2-1 code. The single-figure position on some manufacturers' cards is represented by a V. With double-row punching the effect of an SF position can be achieved by deep-punching the single digits (Fig. 5-6*i*). One code variation handles the problem of coding zeros by using a separate field to indicate the total number of zeros in the number (Fig. 5-6*j*).

Several other coding schemes that also permit relatively rapid serial sorting are available. These schemes, at the cost of using more code positions than the 7-4-2-1 code, incorporate a triangular display on the card in order to simplify the punching and recognition by the user (Fig. 5-6*k*, *l*). Two holes are punched for every digit. The positions to be punched are the ones whose guidelines intersect at the desired digit on the printed display. Serial sorting is performed in the same manner as for the 7-4-2-1 code except that one more hole must be needled. One efficient triangular display scheme uses four code positions with a double-row card (Fig. 5-6*m*). Combinations of shallow (outer row only) and deep (both rows) punching represent the data.

Numbers within a range of values may be recorded by quantizing the number range into a discrete number of intervals, and assigning a code pattern to each interval (Fig. 5-6*n*). Instead of using a range of values, it is possible to use approximate or average values to represent the data, and assign each code position to a particular approximate value (Fig. 5-6*o*).

Direct Alphabetic Information. Figure 5-7 illustrates various notching schemes for representing alphabetic information. The use of a unique hole for each alphabetic character is even more wasteful than it would be for numeric data (see Fig. 5-7*a*). Efficiency can be improved by assigning multiple meanings to each position, using one or more modifier code positions (Fig. 5-7*b*, *c*). One of the most efficient methods is a number representation or count of the particular character, considering the alphabet A to Z to be represented by the numbers 1 to 26. This basic method can be used with single-row or multiple-row punching (Fig. 5-7*d*, *e*). However, the method is more awkward to use because a code table is needed for punching or

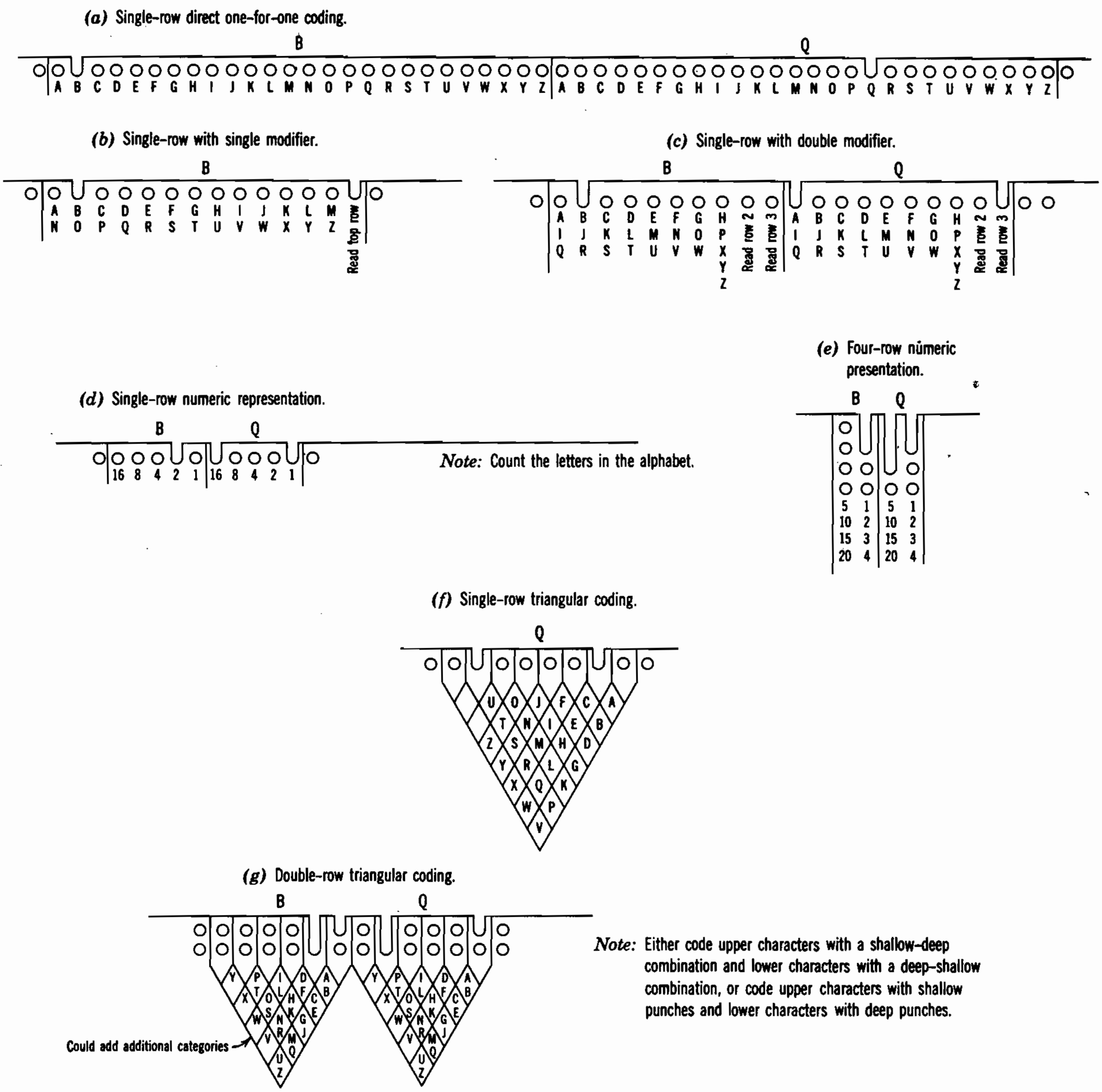


FIG. 5-7 Methods for direct coding of alphabetic data.

reading the holes. As with numeric data, a triangular format can also be used to represent alphabetic data in single- or double-row cards (Fig. 5-7*f*, *g*).

Dates and Times. Figure 5-8 illustrates various methods for representing dates and times by punch patterns. Months or years can be coded directly, with a unique hole for each month or year (Fig. 5-8*a*). The days can be coded by any of the techniques described for coding numeric data. The years can be coded by many of the schemes used to represent numeric data, such as multiple meanings for each hole, unique hole assignment for multiple-row cards, and straight 7-4-2-1, 7-4-2-1-0, or 10-7-4-2-1 numeric representations (see Fig. 5-8*b* to *h*). Coding for the century may be ignored if the card file is restricted to the years within a century.

The months may also be coded by any of the numeric schemes (Fig. 5-8*i*, *j*). One variation of the numeric scheme is to divide the year into quarters, and into months within a quarter (Fig. 5-8*k*). Another variation is to use multiple-row cards to include the coding for the month and the year (Fig. 5-8*l*). Elasped time in months, days, and hours may also be simply encoded (Fig. 5-8*m*).

Names. Figure 5-9 illustrates various methods of representing names by punch patterns. Any of the schemes for coding alphabetic data can be used to completely spell out the names, but this is so clumsy that most users only code some abbreviated form of the name. There are many ways to abbreviate the names (see Chapt. 3), but the most common methods for notched cards are:

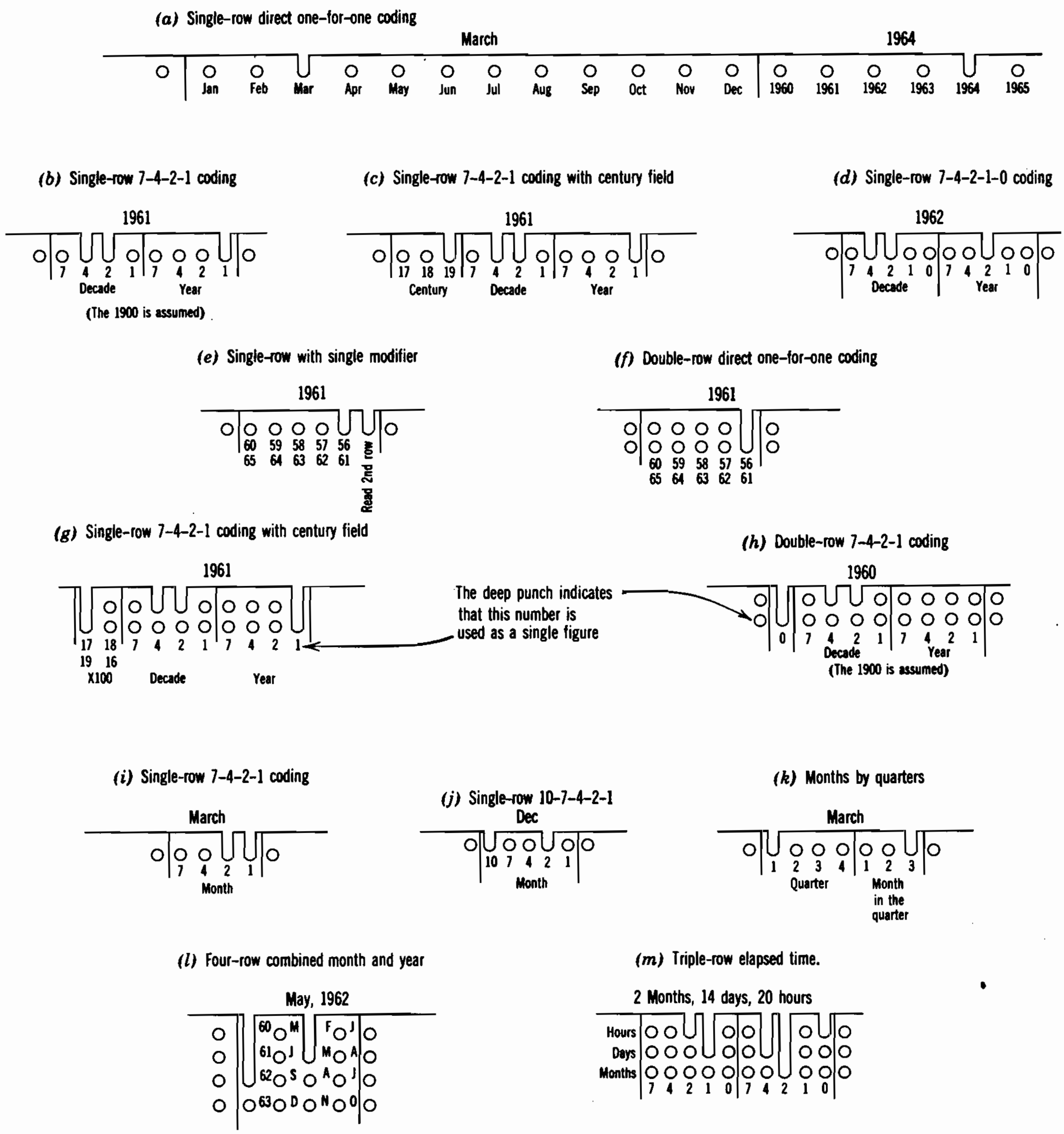


FIG. 5-8 Methods for coding dates or times.

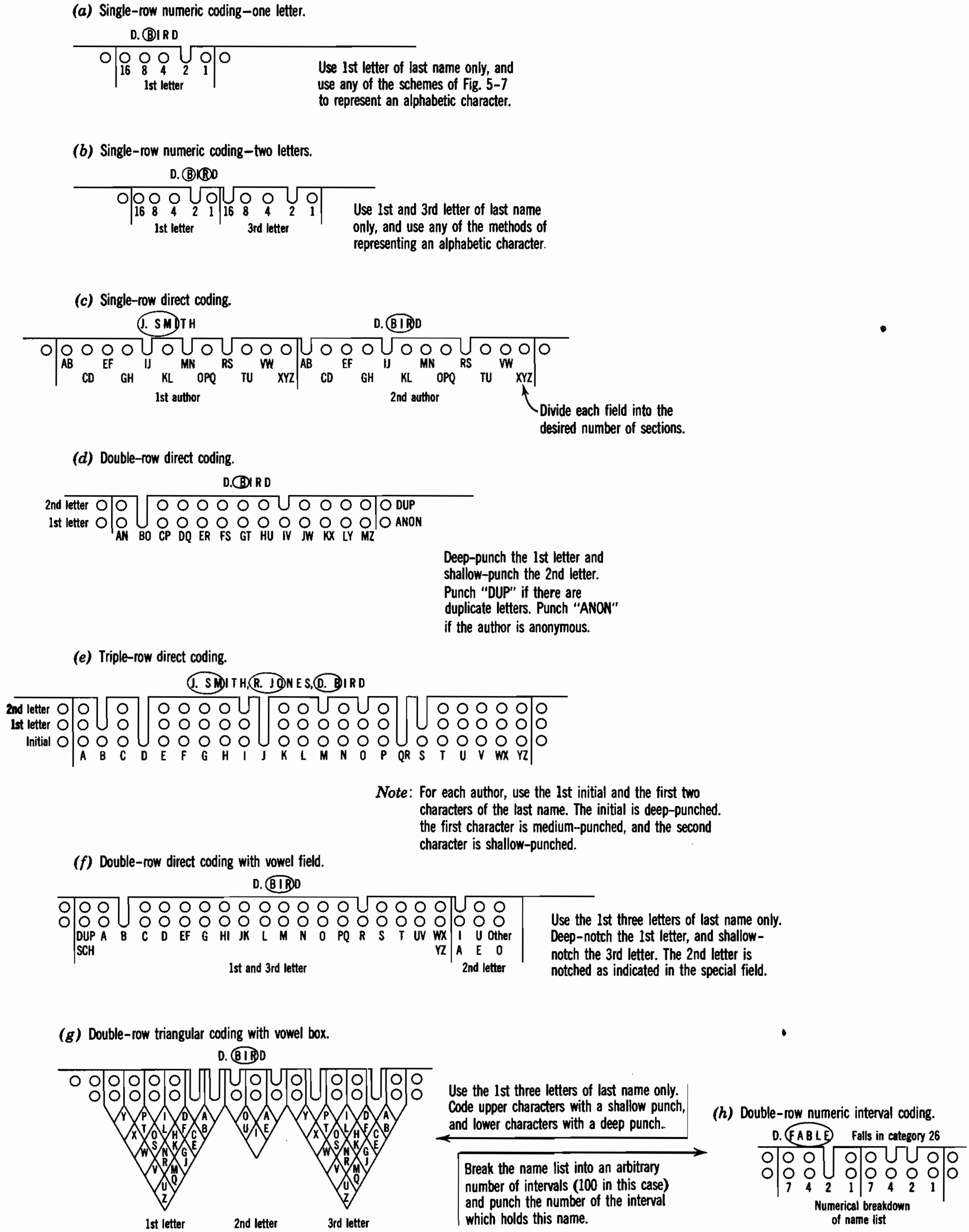


FIG. 5-9 Methods for coding names.

1. Retain only the first initial and the first two letters of the last name.
2. Retain only the first two or three letters of the last name.
3. Retain only the first and third letters of the last name.

Examples of some of the simpler coding methods are given in Fig. 5-9*a*, *b*, c. Figure 5-9c will provide some false drops (e.g., I. MIST or I. SIMPLE as well as J. SMITH) but many fewer than if the initial letter S was punched alone. Direct coding methods using multiple rows are shown in Fig. 5-9*d*, e. For situations (such as several authors) in which two or more names are to be placed on a card, the names are usually put into separate fields, although they may be superimposed onto the same field. As mentioned in Chapt. 3, the second letter of a person's last name is nearly always one of the vowels. Consequently, the second letter does not provide a great deal of discrimination between names or words. Some coding systems have considered this, and have provided a minimum number of coding positions (a, e, i, o, u, and "other") for the second letter (Fig. 5-9*f*, g).

The actual number of holes to be assigned for the alphabetic breakdown is arbitrary, and is usually a compromise among user convenience, the number of available holes, and the required degree of resolution or discrimination between data entries. Coding schemes for names often combine relatively infrequent letters (e.g., XYZ) into the same coding position.

For uniform utilization of the entire coding field, some attempt should be made to assign the characters to the holes in such a way that on the average, each hole position has the same probability of being punched. The frequency of occurrence of the letters, described in Chapt. 3, could be used as a basis for code assignment. It is also possible to use a numeric code such as the one described in Table 3-1 to assign each name to a numbered interval (Fig. 5-9*h*).

Chemical Elements and Compounds. There are many ways in which chemical elements and compounds can be coded, and some of the simpler schemes are shown in Fig. 5-10. Direct coding of the elements can be used (Fig. 5-10*a*), the symbols can be spelled out (Fig. 5-lob), or the atomic number can be given a numeric code (Fig. 5-10*c*). Several comprehensive coding systems have been developed such as the Frear, National Research Coupcil, Dyson, Gruber, and many others.[4] Coding of compounds and elements is relatively specialized, and the code usually reflects the special needs and interests of the individual or organization which is responsible for the file collection. Their interest may be focused upon a narrow subject field such as petroleum products, plastics, and high-temperature metals.

Coding by Categories. Instead of directly punching numeric or alphabetic information into the card, most edge-punched card systems assign specific meanings or categories to each coding position. For example, a fruit cannery that recorded basic data on a card for subsequent statistical studies might assign a separate hole to each of the fruit types (apples, pears, etc.) to be recorded instead of notching out the complete spelling of the particular fruit (see Fig. 5-21). Economy in punching can also be realized by assigning numeric codes to names (e.g., cities and states) or categories of information.

It nearly always proves useful to leave some holes unassigned on the card when the initial format is being planned. Additional requirements for recording data often become apparent after the system has become operational. It is also convenient for the user if the meaning or assignment of each of the holes can be printed directly onto each card. In some cases, it may prove advantageous to print on both sides of the card. This printing can be done by the user with small-scale printing machines such

[4] Frear, D. E. H., "Comprehensive Coding Schemes for Chemical Compounds," Chapt. 22 in *Punched* Cards, 2nd ed., Casey et al., editors, cited previously.

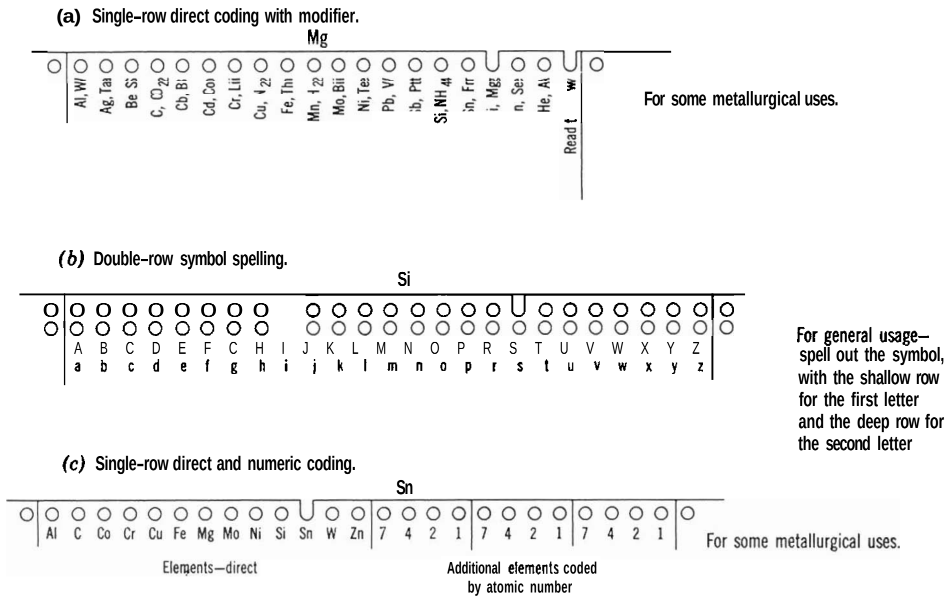


FIG. 5-10 Methods for direct coding of the chemical elements and compounds.

as ditto or multilith systems; however, better print quality is usually obtained if the cards are printed by the card supplier, with letterpress printing equipment.

Superimposed Coding. Superimposed coding, as described earlier, in Chapt. 3, provides a means for more effectively utilizing the storage capacity of a card by superimposing several items of information in the same coding field. Code design techniques are available that allow systems to be developed with relatively compact codes and little "noise" due to superimposed terms. To date, the techniques of superimposed coding have been restricted almost entirely to edge-notched card systems, although some tab card systems have used this method of coding. The most common example of an edge-notched card system with superimposed coding is the Zatocard (see Fig. 3-9), a system developed by Mr. Calvin Mooers, one of the earliest workers in the field of superimposed codes.

The Equipment

Some of the card systems include a small amount of equipment to speed up some of the operations, or to provide additional functions. However, no comprehensive notched card systems exist with capability or performance comparable to the families of punched tabulating card equipment. That is, edge-notched equipment is not available to automatically duplicate decks of notched cards or print a listing of the information notched into the card. Some of the more complex notched card equipment is described below.

Key *Punch.* This unit is used for high-speed or high-volume notching of individual cards. Manual and electric models are available (see Fig. 5-11). The keyboard permits the setting of the coding of an entire card edge before any punching is done, and then it is all performed simultaneously by a single stroke of the motor bar. **A** repeat feature

FIG. 5-11 **Keysort** key punch.

permits the operator to pre-set any column or series of columns of the keyboard to remain in the repeat condition while the keys of all other columns restore after each operation.

Groover. The manual unit simultaneously grooves as many as 50 cards in the same notch position (see Figs. 5-12 and 5-13), while the electric unit simultaneously grooves as many as 150. This feature is valuable when many cards (e.g., a coupon book) or checks must be notched with common information.

Duplicating Punch. This unit reproduces fixed data from a plastic card or Charge-A-Plate onto notched card sets (see Fig. 5-14). This feature is valuable for simultaneously punching several items of fixed data (e.g., name, identification number, department) from a Charge-A-Plate into a card.

Data Punch. This unit has the same capabilities as the duplicating punch, plus the capability for simultaneously imprinting information from the master metal tag or Charge-A-Plate onto the face of the card (see Fig. 5-15). When a card set is used, this information from the Charge-A-Plate is imprinted and punched on all of the pieces in the set, thus generating a paper copy for other record-keeping operations. Maximum coding capacity is 8 digits on one edge of the card. The plates may be used for imprinting without notching.

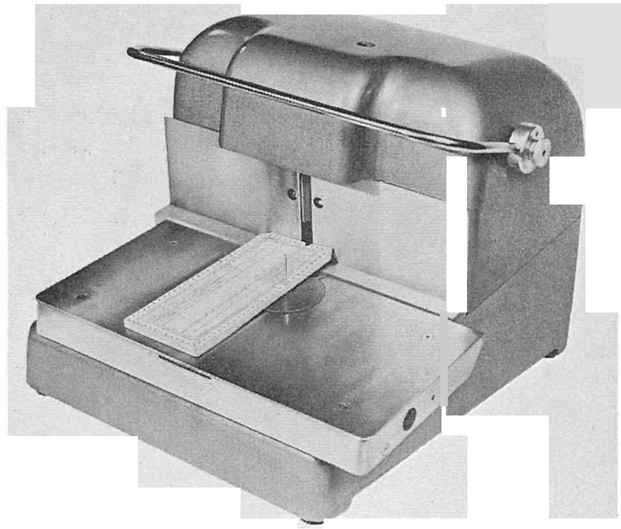


FIG. 5-13 **Keysort** electric groover.

Tabulating Punch. The tabulating punch unit does not read or interpret the punching on the edge of the card (see Fig. 5-16). However, it can punch numerical data into the interior of the card in a

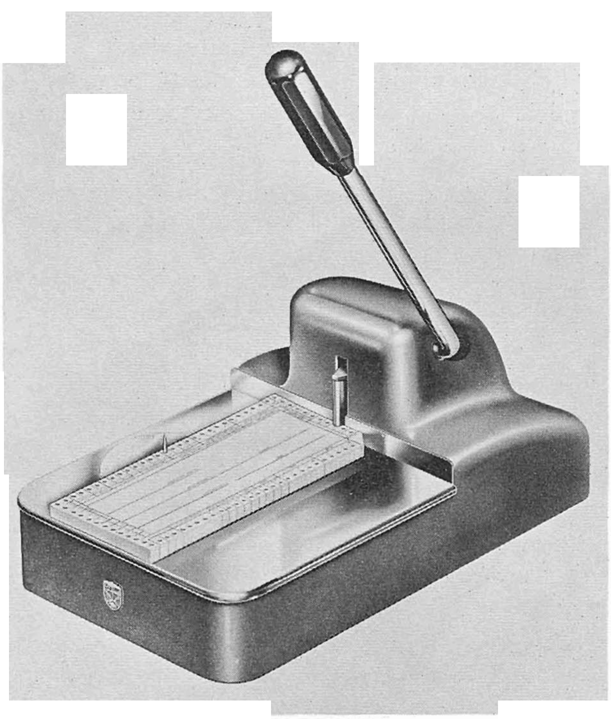


FIG. 5-12 **Keysort** card groover.

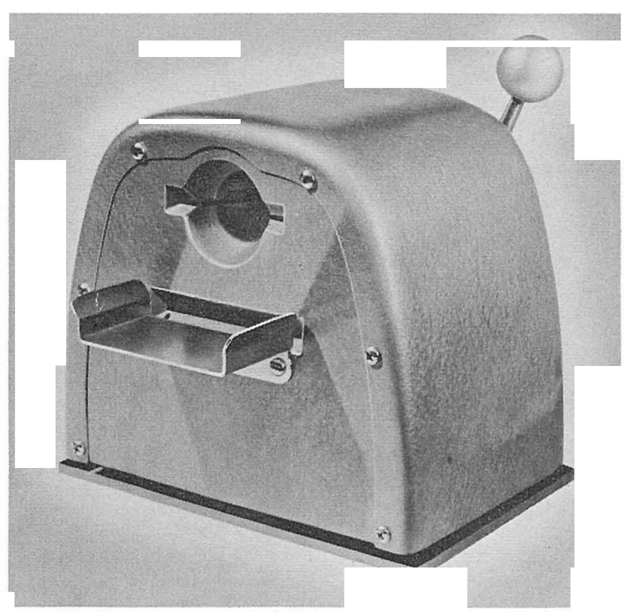


FIG. 5-14 **Keysort** duplicating punch.

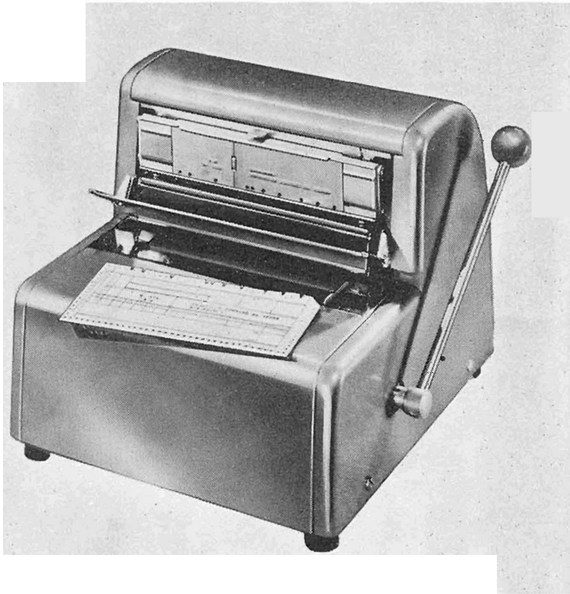


FIG. 5-15 Keysort data punch.

form similar to paper tape coding, and it can later read such numerical data in order to add, subtract, or print them with the information from the cards. The unit can also reproduce a card while automatically printing and totaling the information previously punched into the card's interior by the tabulating punch.

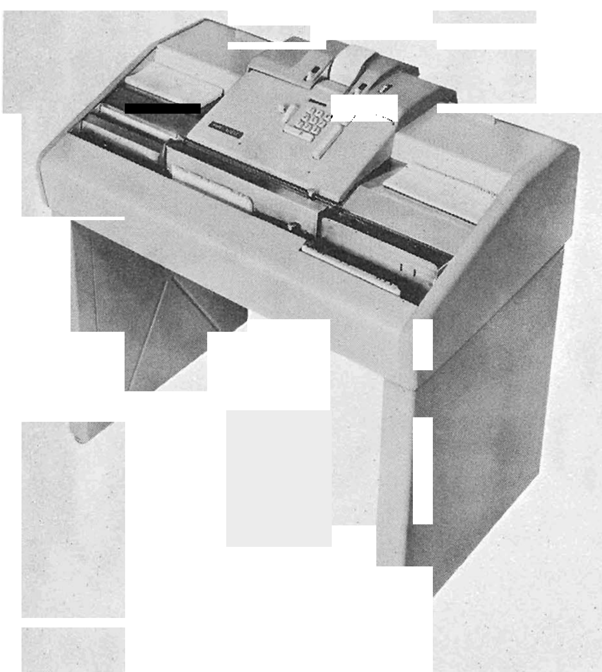


FIG. 5-16 Keysort tabulating punch.

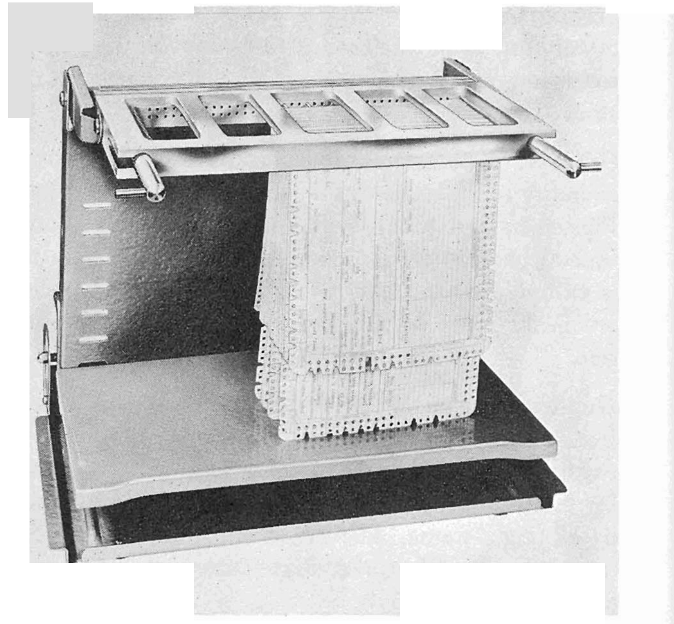


FIG. 5-17 Keysort selector.

Keysort Selector. The Keysort selector is a mechanical framework that can be used to speed up the selection operations, especially where a relatively large number of hole positions must be needled (see Fig. 5-17). The selector is especially

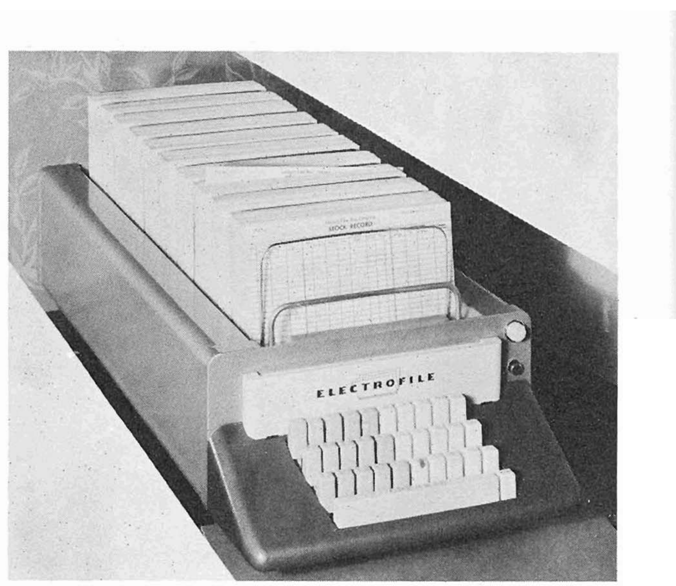


FIG. 5-18 Electrofile card system—single unit.

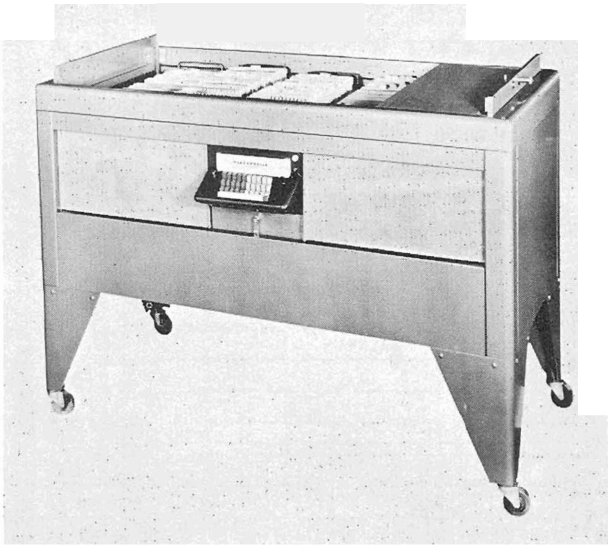


FIG. 5-19 Electrofile card system—multiple unit.

useful for doing multiple needle sorts without disturbing the file order. One entire edge, either single-row or double-row coding, is sorted in one operation. Approximately 250 cards are handled in each selection operation, and the file order of the selected and rejected cards remains undisturbed.

Zatocard Selector. The Zatocard Selector is a device which permits rapid searching of decks of Zatocards (see Fig. 5-3). This is especially useful when a large number of hole positions are to be needled. The separation of the selected cards is speeded up somewhat by a vibrator, which shakes the cards loose from each other.

Electrofile. The Electrofile stores and searches files (see Figs. 5-18 and 5-19) of specially constructed edge-punched cards. The cards have a toothed magnetic metal strip at the bottom which is used for coding the card (see Fig. 5-20). After the search specification has been typed on the Electrofile keyboard, the device selects the pertinent cards from the file without any manual card handling. The main Electrofile search mechanism may operate with a single tray of cards or with multiple trays of cards. When the search mechanism's keyboard buttons are depressed, corresponding selector rods are pushed up between the teeth in the bottom of the cards, so that cards that have not been punched with the proper code are restrained from moving to one side when an electromagnetic field is applied. Properly punched cards are pulled to one side of the deck by the magnetic field, and then tilted upward on one corner so that the operator can easily observe and extract them. Because of the method of selection, the cards may be filed in random order in the trays. The cards are notched by the same keyboard that is used for the search.[5]

Applications

Reporting and Charging System. In many cases a document must be quickly generated to record some transaction or event. This document may not require an extensive amount of sorting, and does not have to remain an active file item long. Thus, such a document can be used as a temporary note or scratch pad, rather than as a permanent record. The following examples describe several different ways in which notched cards have been used as transaction records in reporting or charging systems.

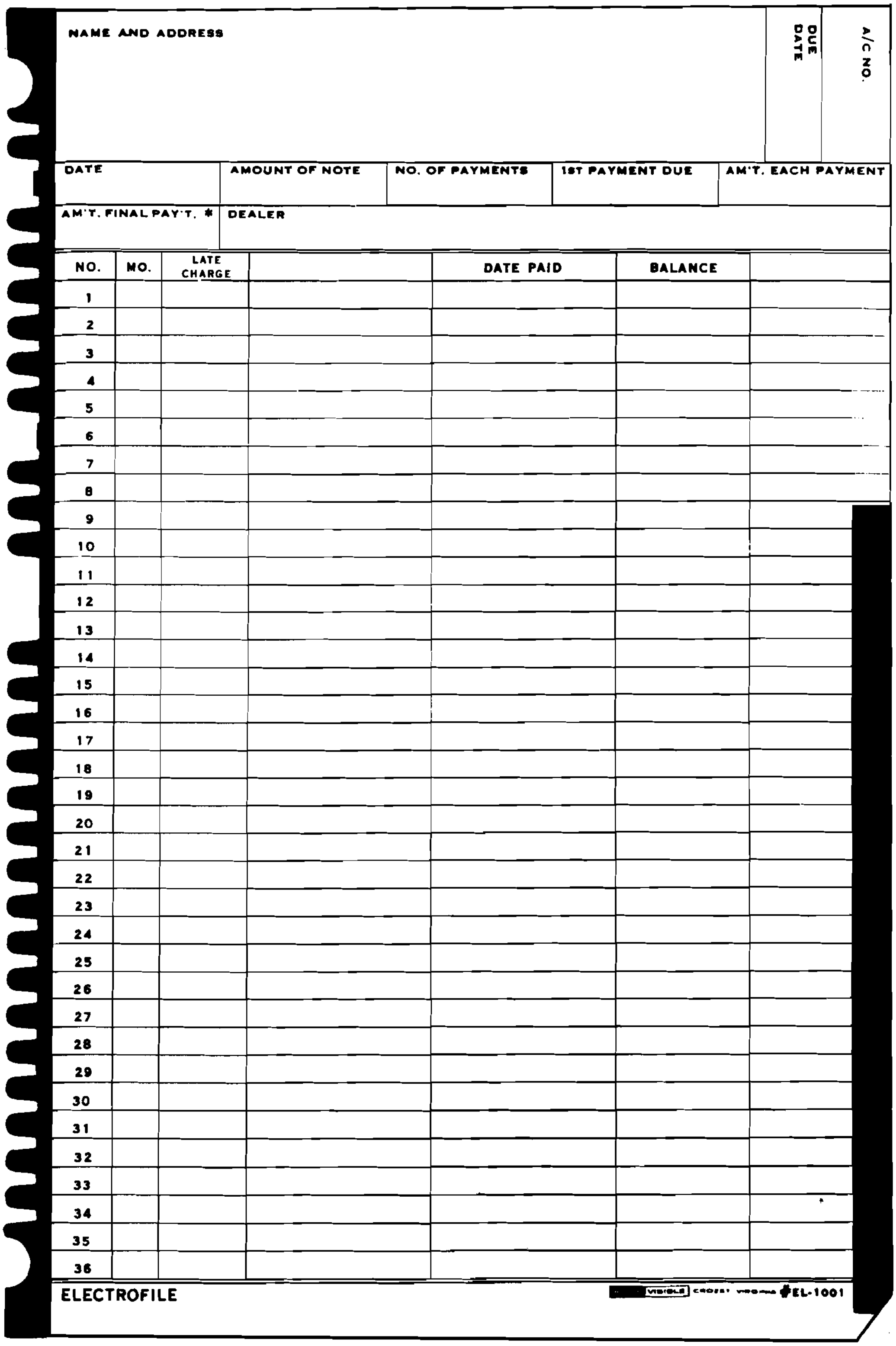
TIME CARDS. Figure 5-21 shows a card used by a large fruit and vegetable cannery to record the direct labor charges of the hourly produc ion workers. Notched cards have been used for this purpose in several other installations.[6] The card shown is a good example of category coding, in which most of the individual holes in the card stand for a specific category such as the day, shift, department, and type of fruit or vegetable worked on. The employee number is the only item numerically coded into the card. Twenty-four holes are numbered and reserved for spares and special uses such as special operations in each department. This card is notched by the foreman for each of his workers at the end of the shift, and used primarily to generate employee payroll data. It is also used for cost accounting studies to determine the production labor costs to process each of the fruit or vegetable types.

CHARGE TICKETS. Figure 5-22 illustrates a hospital patient's charge ticket for laboratory services. This type of ticket is in fairly widespread use for hospital accounting, with format variations to satisfy the

[5] Anon., "Automatic Loan File," *Auditgram,* December 1955.

Anon., "Dealers Install Automatic System for Degree Day Deliveries," *Fuel Oil News,* May 1955.

[6] Anon., "Keycard Simplifies Labor Distribution at Norge," *Management and Business Automation,* Vol. 2, No. 4, pp. 28-30 (October 1959).



NAME AND ADDRESS

DUE DATE

A/C NO.

DATE	AMOUNT OF NOTE	NO. OF PAYMENTS	1ST PAYMENT DUE	AM'T. EACH PAYMENT

AM'T. FINAL PAY'T. #	DEALER

NO.	NO.	LATE CHARGE		DATE PAID	BALANCE	
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
11						
12						
13						
14						
15						
16						
17						
18						
19						
20						
21						
22						
23						
24						
25						
26						
27						
28						
29						
30						
31						
32						
33						
34						
35						
36						

ELECTROFILE

VISIBLE CROZET VIRGINIA #EL-1001

FIG. 5-20 **Electrofile** card.

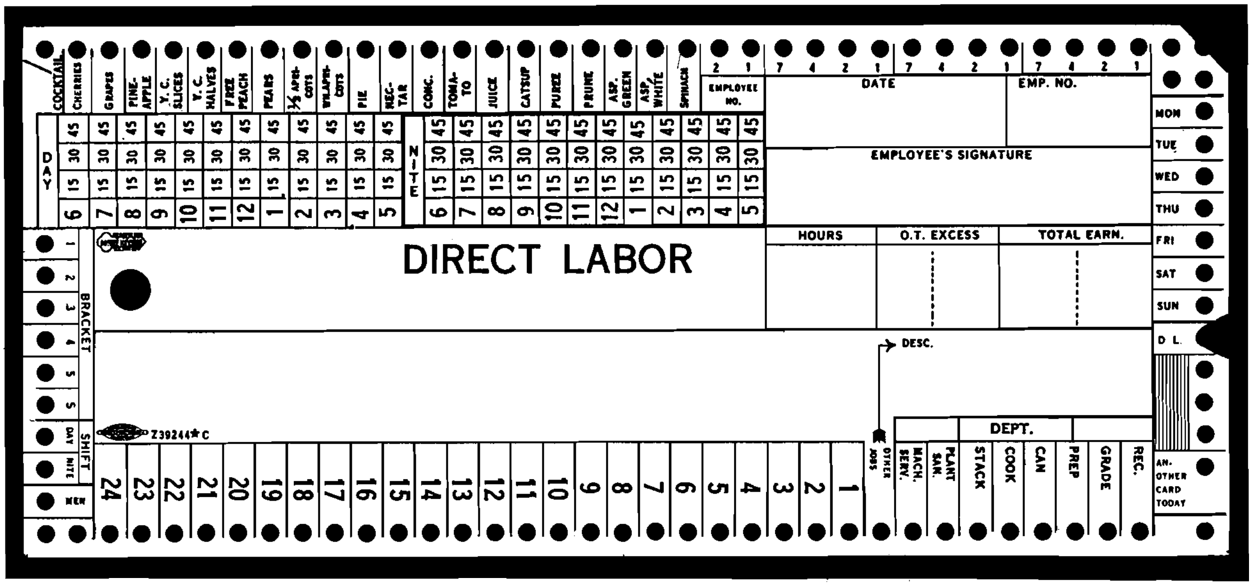


Fig. 5-21 Time card.

local requirements. The ticket set shown consists of three carbon-backed papers (copies for the nursing section, patient's chart, and laboratory) and a sorting card (for the business office). For a situation in which several documents must be generated, each with some common data, an imprinting punch machine (Fig. 5-15) can be used to print and punch the common data onto the charge ticket. A metal plate similar to a Charge-A-Plate is made for each incoming patient, and is used to label all the patient's charges and reports during the hospital stay. This includes the registration forms, the patient's medical record, and any request for medicines or special services. The patient's plate contains fixed information such as the patient's name and number, doctor's name, type of accommodation, and other pertinent data. The nursing station in the hospital that is responsible for this patient also has

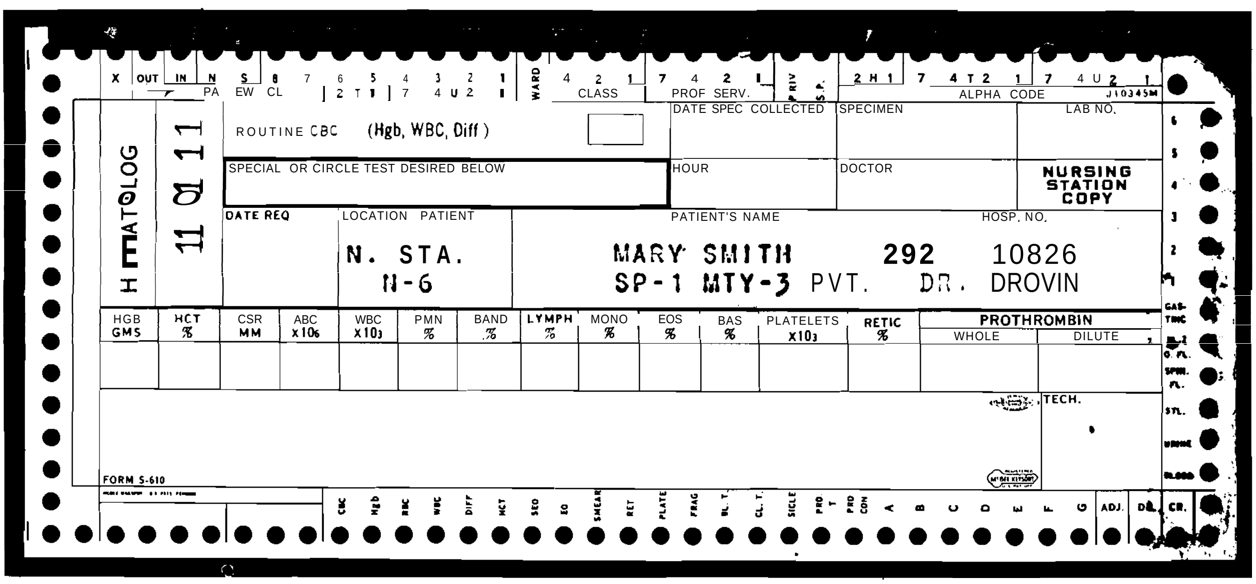


FIG. 5-22 Hospital charge ticket.

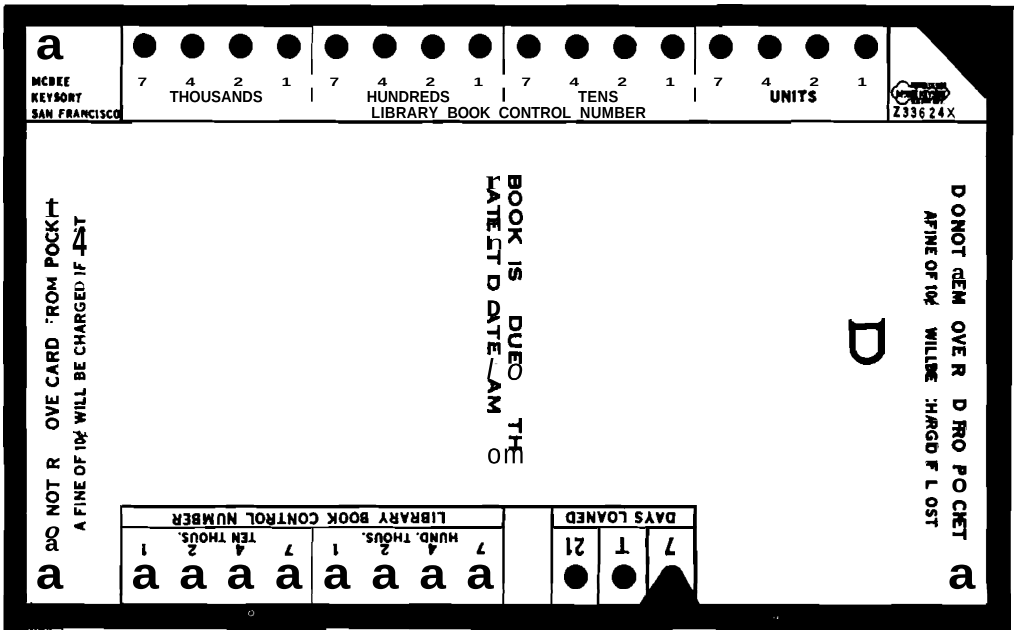


FIG. 5-23 Library charge card.

an imprinting plate to describe the physical location of the patient (Nursing Station N-6 in the example). To generate a charge ticket, the patient's plate and the nursing station plate are jointly inserted in the punch, and the basic charge ticket is notched as shown in the example. Additional details and variable data are then filled in as required.

LIBRARY CHARGE-OUT SYSTEM. Very simple notched cards systems (see Fig. **5-23)** have been used in many libraries to record and control the charge-out and circulation of books and other material. Mrs.

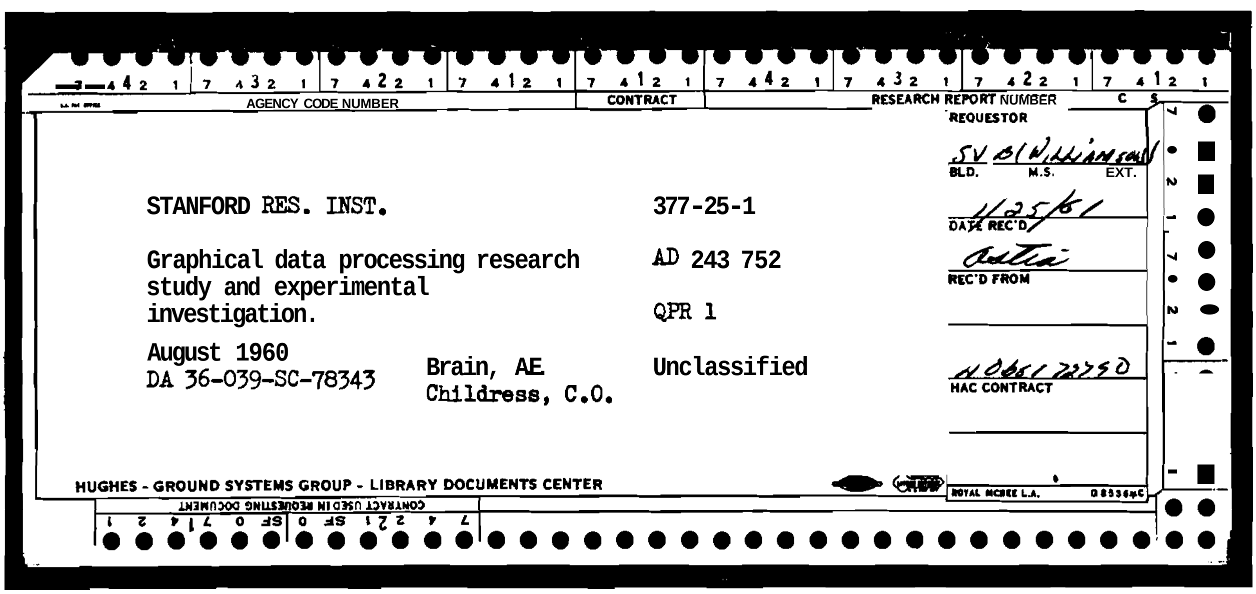


FIG. 5-24 Spirit stencil card for library catalog cards and circulation control.

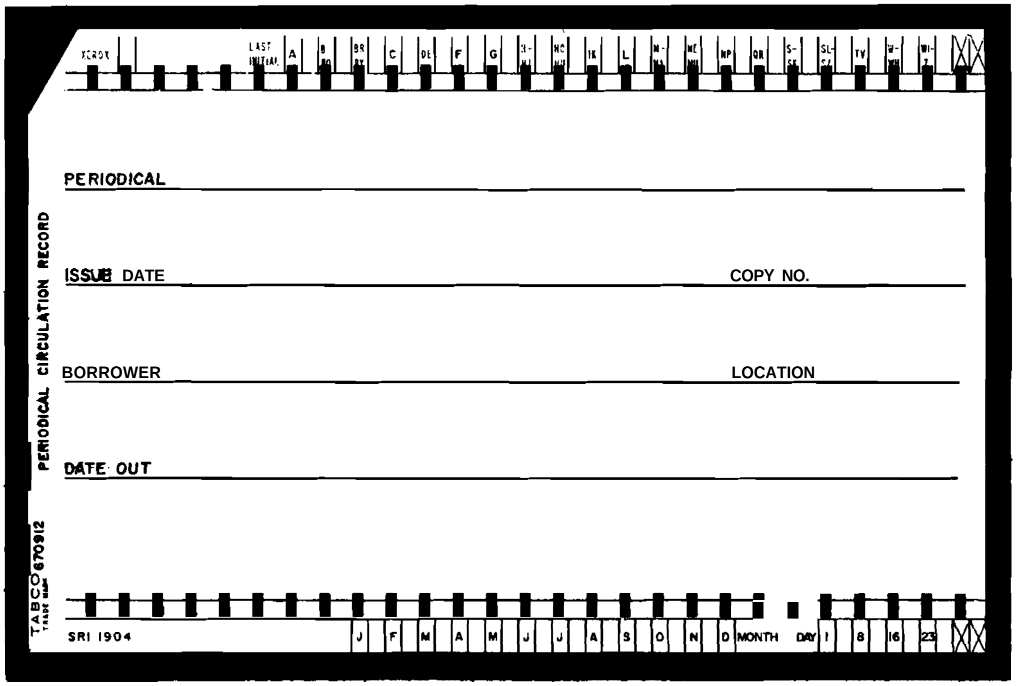


FIG. 5-25 IBM card used as an edge-notched card for library circulation control.

Mary Lou Haire at the Hughes Aircraft Company has developed and installed a clever notched card system for a document center that serves many library functions besides circulation control.[7] The key element in this system is a notched card which contains the complete citation of a document or book typed onto a spirit stencil insert. The stencil card is prepared as soon as the document enters the system, and is subsequently used to put a heading on the indexing worksheet, prepare an accessions list, prepare the required number of 3-by-5-inch catalog cards, and prepare sets of charge-out cards (an original charge card, a first overdue notice, and a second overdue notice). The stencil card (see Fig. 5-24) may be sorted by corporate author, contract, report number, and the contract for which the document was requested. The charge cards may be sorted by the borrower's badge number and due date.

Another example of a circulation control card is given in Fig. 5-25. This is a rather special type of card in that it is a regular and pre-printed 80-column IBM card which has been pre-punched with the character I in alternate columns, in order to provide a row of holes at the top and bottom of the card that could be used for notching and needling in the same manner as the other edge-notched cards, and then trimmed at the print shop to make a shorter card. Special rectangular needles are available from IBM and other tab card equipment suppliers, that can be used to needle the cards. The main reason for using an IBM card in a manual system is an economic one—the total cost of the pre-printed, pre-punched IBM card is appreciably less than the cost of an equivalent card from one of the edge-notched card suppliers. The use of IBM cards for this purpose was first suggested in the literature by Ohlman.[8]

REQUISITION FORMS. Cards or card sets are often used as parts or material requisition forms in production operations to easily assemble all pertinent charges to the proper project or cost center. The example shown in Fig. 5-26 is used by a large manufacturing organization to record miscellaneous material requisitions. Instead of writing in the parts

[7] Banks, D. H., "Planning the New Library: Hughes Aircraft Company's Ground Systems Group Technical Library," Special Libraries, Vol. 52, No. 4, pp. 191-197 (April 1961).

[8] Ohlman, H., "The Low-Cost Production of Marginal-Punched Cards on Accounting Machines," American Documentation, Vol. 8, No. 2, pp. 123-126 (April 1957).

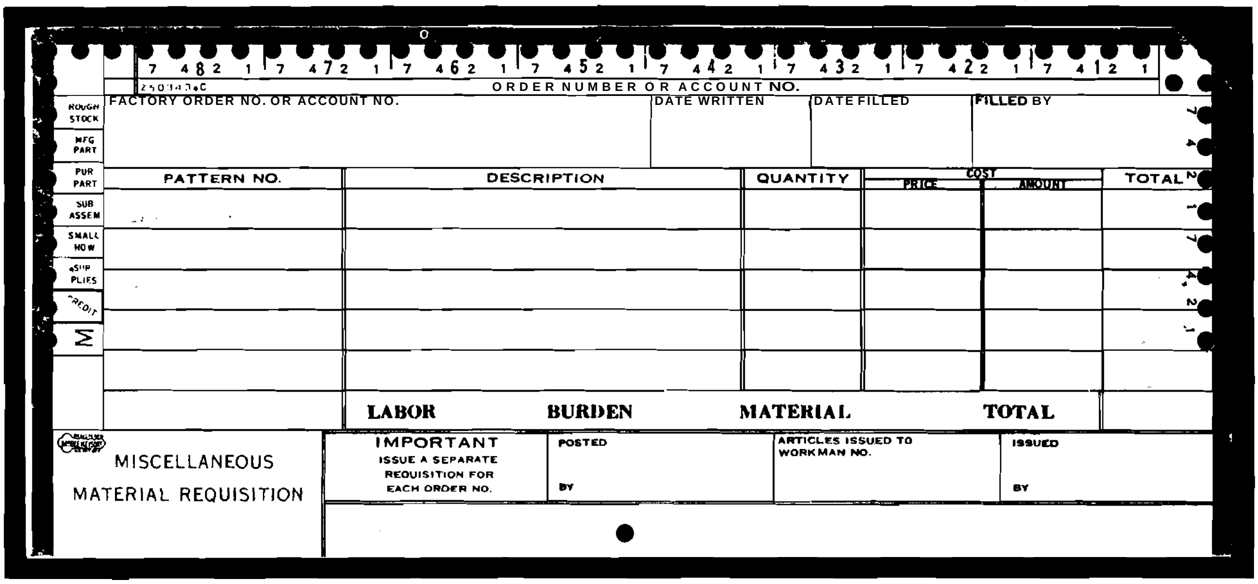


FIG. 5-26 Requisition card.

description, it may be desirable in some instances to code and notch the description of the item which is being ordered, so that the cards can also be used for inventory control purposes.

Technical Data Storage. Notched cards are often used for situations in which a relatively large amount of technical data must be correlated, or searched. Such data files are usually of a very specialized nature, intended to satisfy the needs of one particular user, although in some cases such files may be useful to other organizations. Several representative applications are given below.

TRANSISTOR INDEX. There are so many electronic components (e.g., tubes, transistors, resistors) avail-

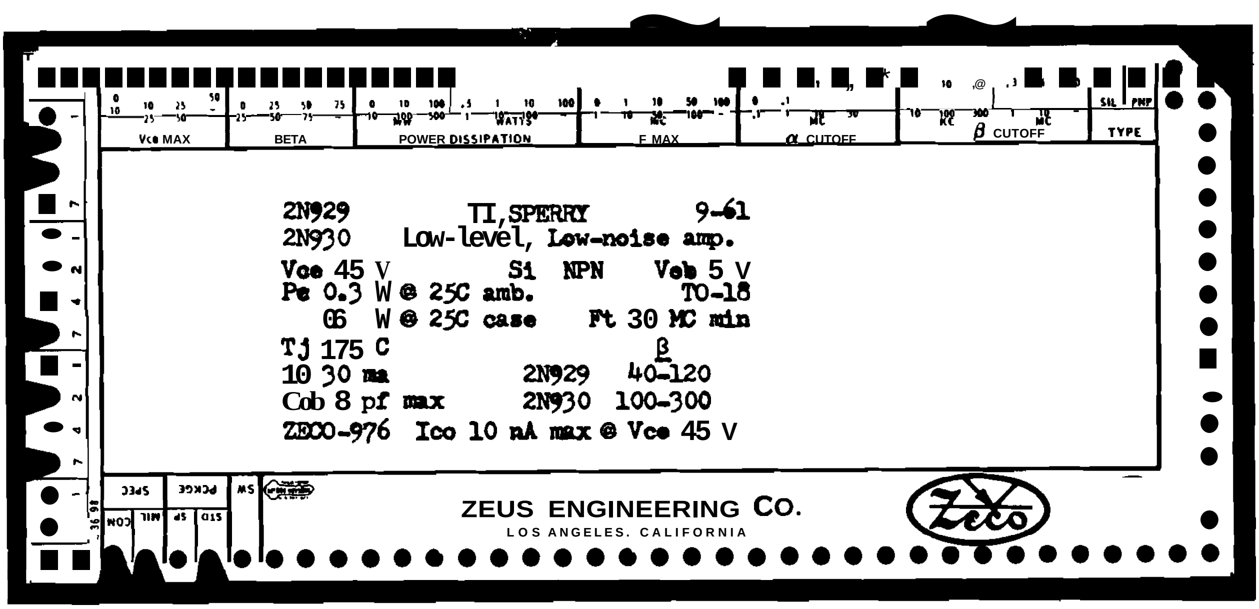


FIG. 5-27 Transistor data card.

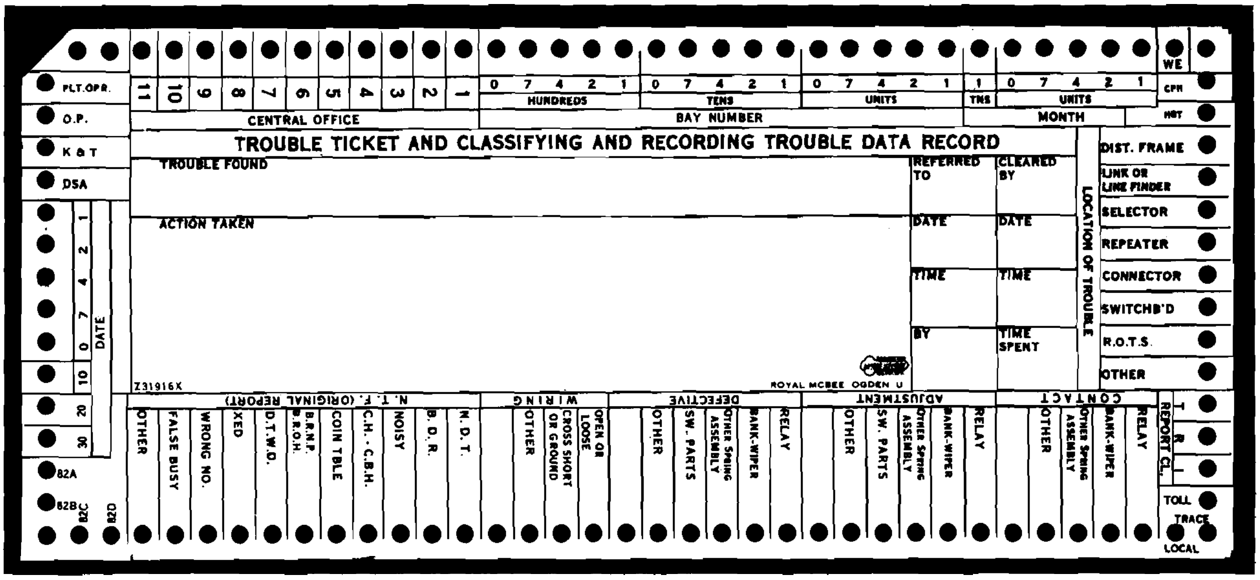


FIG. 5-28 Trouble ticket.

able for use by design engineers that many engineering organizations have established or subscribed to card files on components and devices. Several commercial firms sell subscriptions to such card files and tabulated listings of components characteristics. One example of such a service is the Zeco Transistor Index.[9] A copy of the type of card which they provide in their file service for transistor characteristics is illustrated in Fig. 5-27. Each card describes a particular transistor or family of transistors by means of both notched sorting slots and printed data on the card. The transistor information is printed on the face of the card with an office duplicating machine.

PERFORMANCE OR TEST DATA. Notched cards are useful for the collection and analysis of performance data such as the maintenance and parts replacement records of radios, computer systems, or automobiles. These data are often collected in order to statistically analyze the malfunctions and required replacements. One example of the use of a card for recording performance or trouble data is given in Fig. 5-28. This particular card is used by a telephone company for statistical studies of the location and types of trouble in the telephone system equipment.

[9] Distributed by the Zeus Engineering Co., Inc., 625 Kingsley Drive, Suite 1, Los Angeles 5, California.

MATERIAL COMPOSITION OR CHARACTERISTICS. As mentioned earlier, techniques are available for describing chemical compounds with a general and systematic coding scheme. But the majority of card files for chemical data seem to use category coding, with a particular compound or element assigned to holes on a one-for-one basis. Cards have been designed in this way for special chemical data, geological or mineral data, metallurgical data, and many other special types of data.[10] Figure 5-29 is an example of a card which records the description and measurements of some of the properties of special chemical compounds. Additional descriptive information is recorded on the reverse side of the card.

PART SPECIFICATIONS. Some organizations may have a large number of parts in their product line. In such cases, the user may want to be able to select a particular product by one of a number of characteristics of that product. In this respect, this application is similar to the use of cards in the transistor file that was previously described. One example of such a part description card is given in Fig. 5-30.

[10] Anon., "Mass Spectra Data on Keysort Cards," Chemical *Engineering* News, December 29, 1952, p. 5466.

Kuentzel, L. E., "Punched Cards as Aids to Qualitative Chemical Analysis by Spectral Methods," Chapt. 9 in Punched Cards, 2nd ed., Casey et al., editors, cited previously.

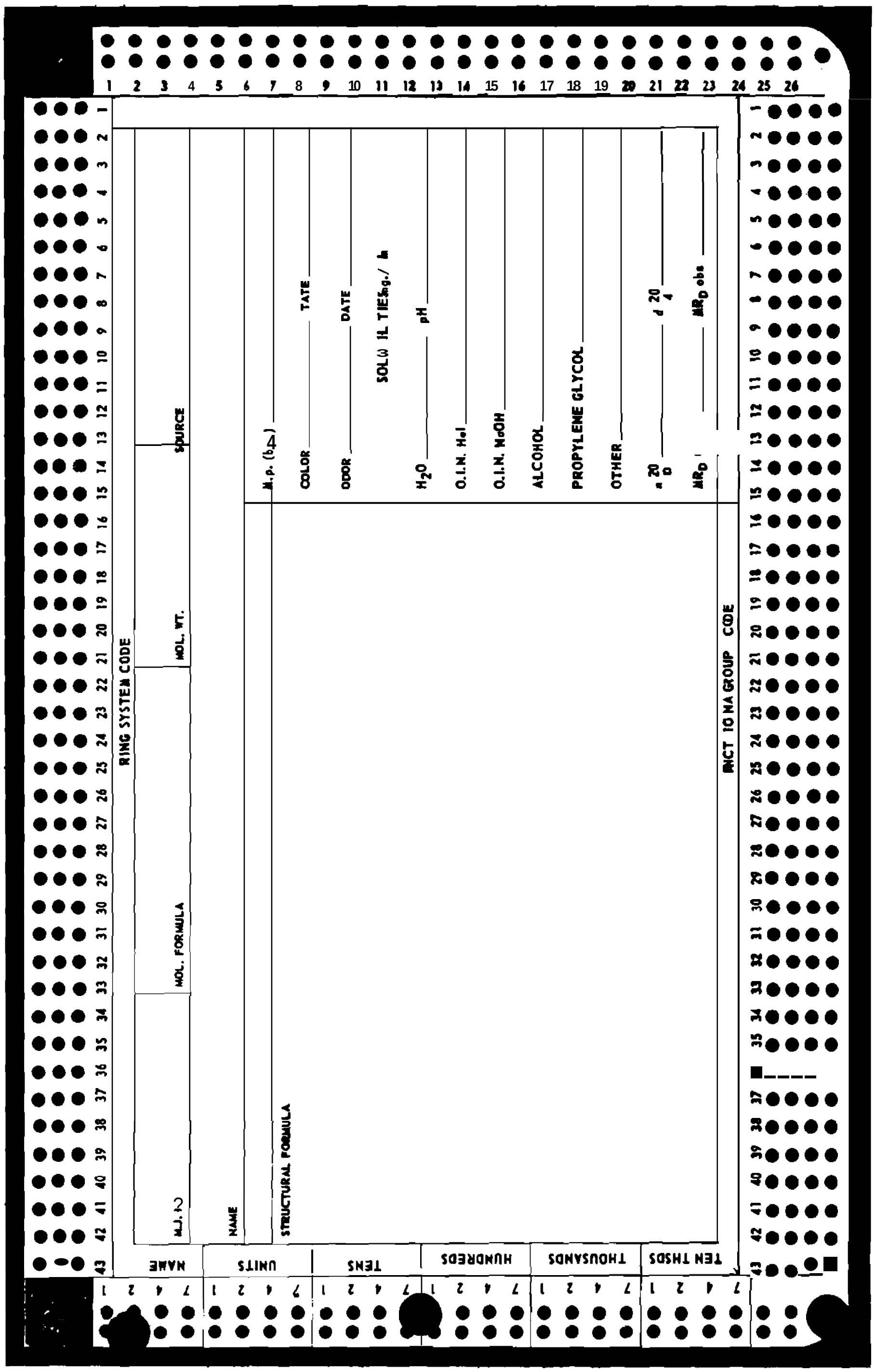


FIG. 5-29 Chemical compound file card.

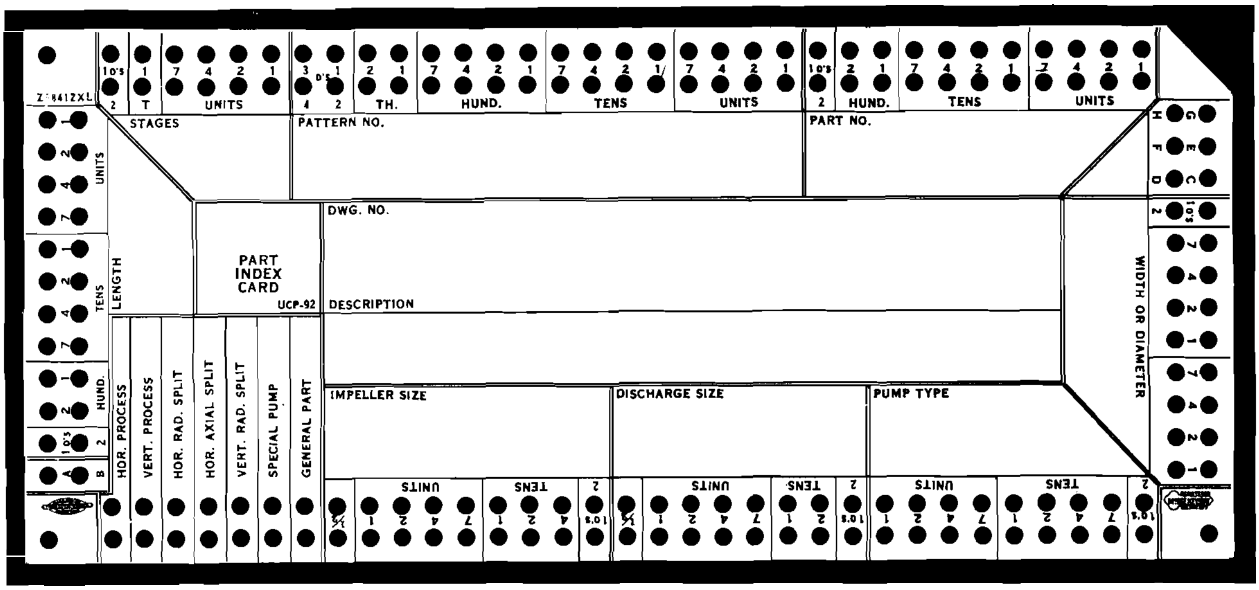


FIG. 5-30 Part index card.

Scheduling Systems. The examples below are given to illustrate how the cards may be used to assist the planning and scheduling of personnel and facilities.

CLASSROOM SCHEDULING. Figure 5-31 describes a card set which is used by a high school during the student's class registration period in order to assist the scheduling of classes and the allocation of classrooms and teachers.[11] After the student's proposed program content has been agreed to by the parents and the counselor, the pertinent information is notched into the card. The three or more parts of the card set are then distributed to the parent, counselor, principal's office and other parties as required. The heavy card in each set is often used as a master student locator card in the principal's office. With this information in each of the students' cards, the complete collection of cards can then be sorted to determine the advance enrollment in each of the classes. This will quickly provide the necessary information for scheduling the classes and facilities.

PRODUCTION SCHEDULING. Production work centers can be scheduled and allocated in a manner similar to that described for classroom scheduling.[12] Data to show which machines, processes, or departments are affected by a particular order are notched in the cards, and then the cards are sorted to observe the predicted workloads for each work center. The pertinent work center and operation number are notched into the card (see Fig. 5-32). However, the scheduling details are not notched, and instead, space is provided on the card to manually record the required scheduling. This particular example (called a *Load Schedule Card)* is one of a group of 5 different cards used as part of a production control system for a large equipment manufacturer organization. The other 4 cards are used for direct labor charges, material requisition, miscellaneous material requisition, and summary data. This particular manufacturer currently uses about 400,000 cards each year for the production scheduling and cost accounting of a division which has approximately 400 shop workers, and distributes 400 to 500 production orders per day to the shop.

Personnel Records. Cards are often used to store background data on an organization's personnel, in order to easily locate people with particular skills or backgrounds, or to provide a means for determining the total background or skills of the entire per-

[11] Matulis, A. S., *Case History #978. How Taylor Center High School Achieved Homogeneous Student Ability Grouping-and Many Other Benefits, with Low Cost Keysort Notched Card System,* Brochure S-570 of the Royal McBee Corp., Port Chester, New York, 1960.

[12] Crause, S. J., "How a Simple Control System Cut Production Costs," *Office Equipment and Methods,* Vol. 2, No. 1 (January 1956).

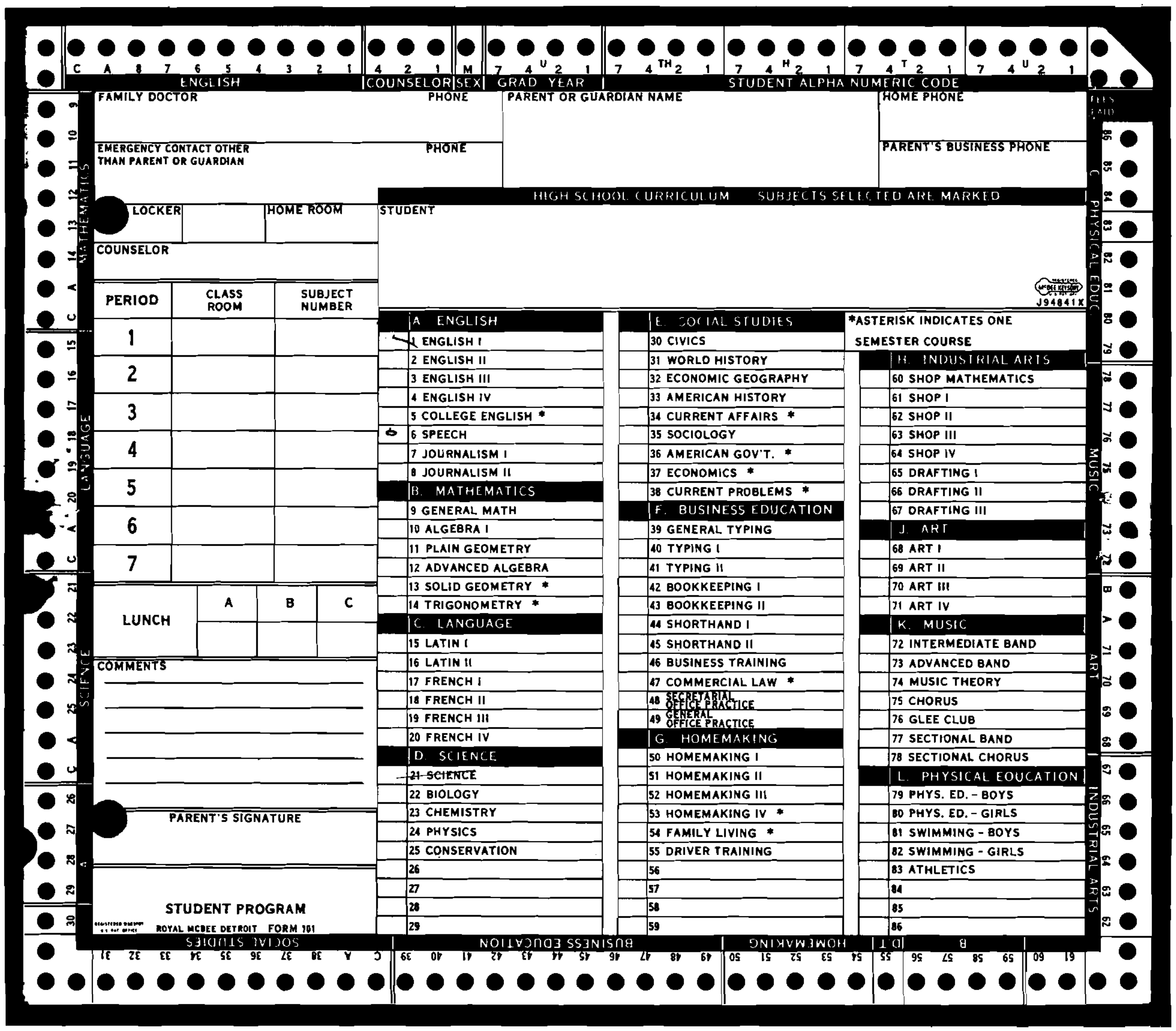


FIG. 531 Student program card.

sonnel roster. Information such as additional education or academic degrees, and job skills, can be added to the card at some later date if necessary. However, information that might be altered, such as salary, marital status, present department, and subject interests, presents more difficulty. Figure 5-33 is an example of one type of personnel card. The reverse side of this card is used to record changes, re-assignments, and factory courses completed.

Notched card systems are also used to describe a criminal's background and reports of arrests.[13]

[13] McPherson, L. A., "Modernized Records System for the Small Department," The Police Chief, February, 1958.

Performance ratings or reviews such as the school report card are another type of personnel record. Figure 5-34 shows a punched card set that can be used to conveniently provide the documentation associated with the reporting of students' grades. At the beginning of the term, sets of cards are prepared for each student, with one set for each class or subject to be taken. Fixed data such as the identification numbers for the student, instructor, counselor, and subject are punched into the cards at this time. Each of the papers in the set is carbon-backed in such a way that information entered on the top sheet is always reproduced on the sheets below it. In the example shown, information about the stu-

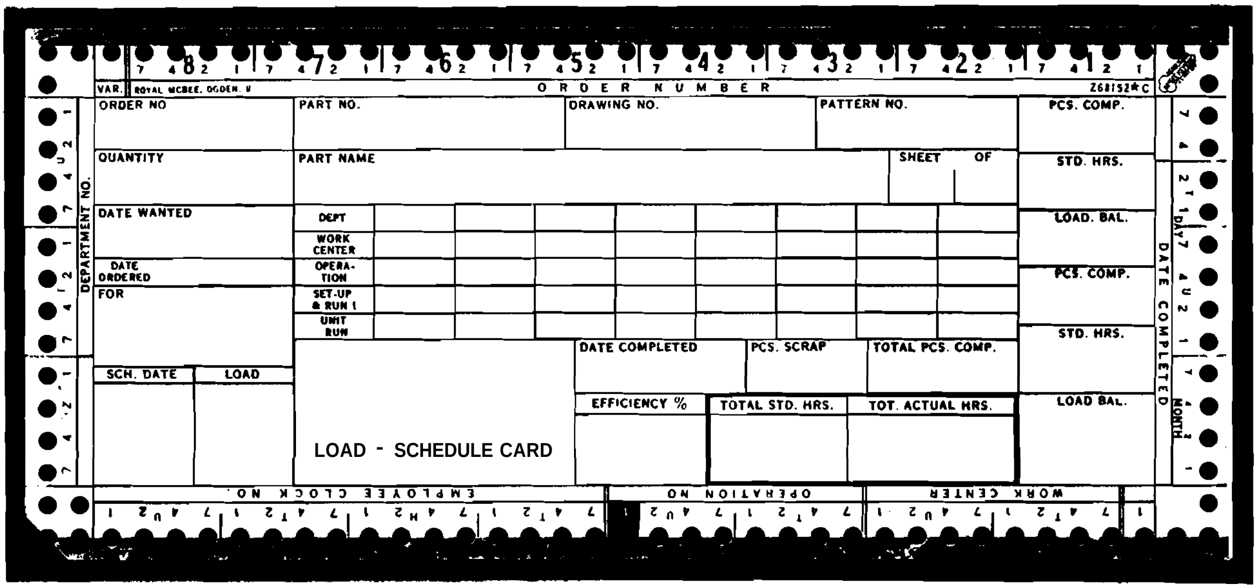


FIG. 5-32 Production scheduling card.

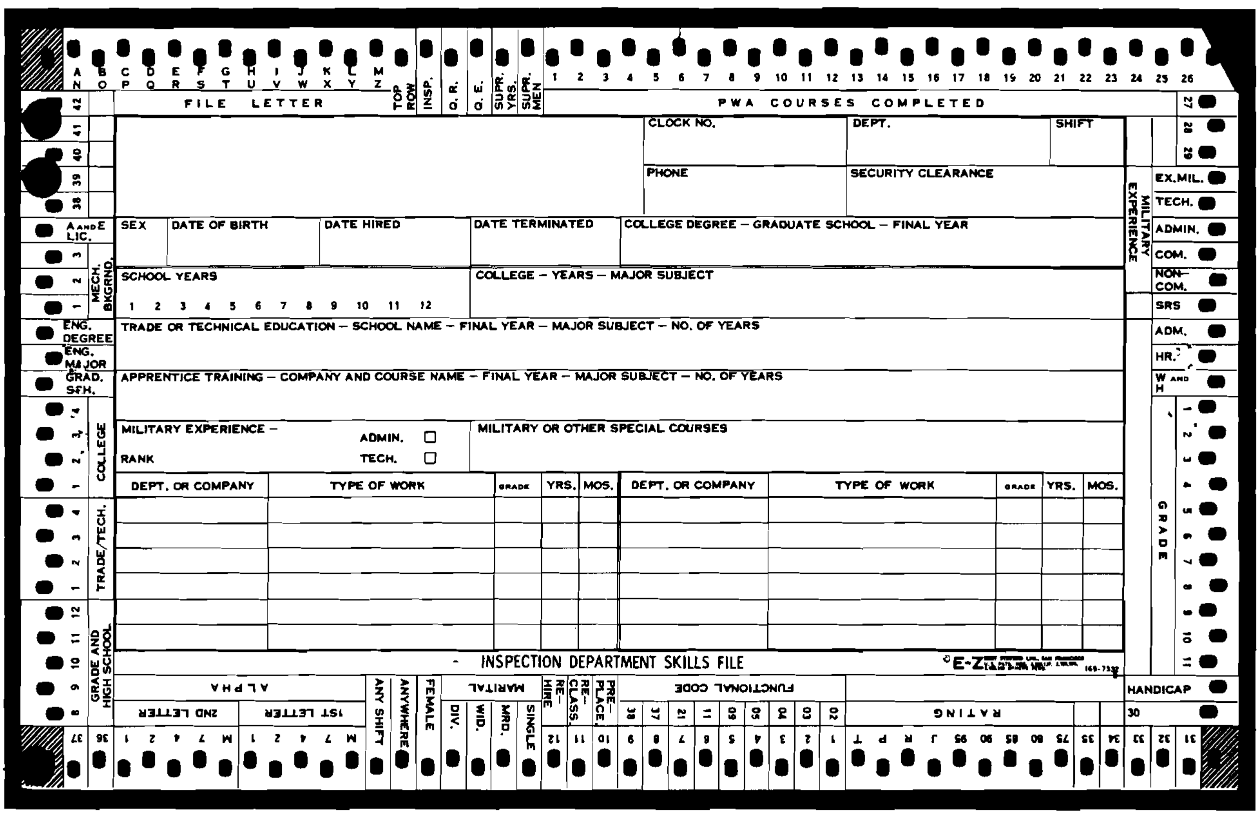


FIG. 5-33 Personnel record.

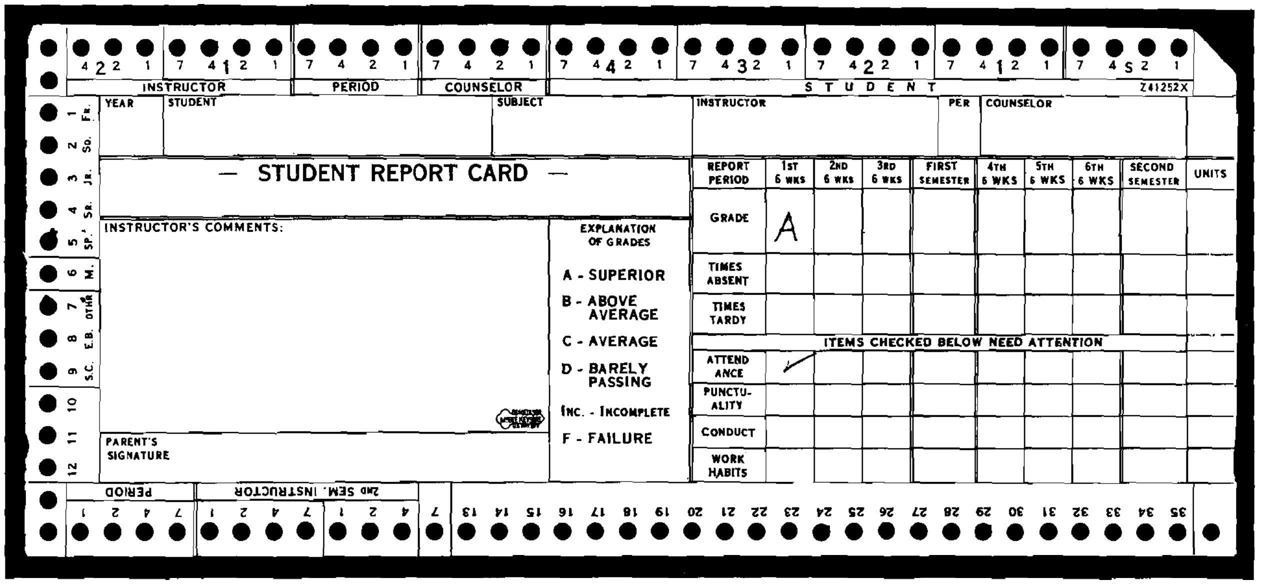


FIG. 5-34 Student report cord.

dent, instructor, course, and performance is transferred through to all the copies beneath the top sheet. For the first report, the grade and comments are written into the appropriate locations on the top sheet, and automatically reproduced by the carbon backing onto all of the remaining sheets. The top sheet then serves as the first report card to be taken home for the parents' signature. The next sheet serves as the report card for the second period, and so on. All the subsequent report cards will contain the grades of the previous report periods. The final card in the set is used by the administration for the student's cumulative records.

Bibliographies. There has been a considerable amount of interest in the use of notched cards for the preparation of bibliographies and reference files.[14] Many such applications have been described in the literature, with many variations in the method of coding the subject, the authors, and the references. Most of the reported systems were developed to satisfy the needs of a particular individual or organization. Names and dates may be coded in the manner described earlier in this chapter. Journal titles may be coded by the direct nu-

[14] Westbrook, J. H., and L. H. De Wald, "A Modified Punch Card Filing System for Metallurgical Literature," *Metal Progress,* Vol. 54, No. 3, pp. 324–327 (September 1948).

Milne, L. J., and M. Milne, "Foresight and Hindsight on a Punch-Card Bibliography," *American Documentation,* Vol. 10, No. 1, pp. 78–84 (January 1959).

Mathay, W. L., and R. B. Hoxeng, "A Classification and Filing System for Corrosion Literature," *Corrosion,* Vol. 12, No. 11, pp. 588t–592t (November 1956).

Staats, H. N., "Data Extraction in Nondestructive Testing," *Journal of the Society for Nondestructive Testing,* Vol. 15, No. 1, pp. 14-46 (January–February 1957).

Cox, G. C., C. F. Bailey, and R. S. Casey, "Punch Cards for a Chemical Bibliography," *Chemical and Engineering News,* Vol. 23, No. 18, pp. 1623–1626 (September 25, 1945).

Hyslop, M. R., and A. Wassenberg, "An International Classification and Punched-Card Filing System for Metallurgical Literature," Chapt. 5 in *Punched Cards,* 2nd ed., R. S. Casey et al., editors, cited previously.

Hyslop, M. R., "An Edge-Notched Punched Card System for Metallurgy," Chapt. 10 in *Information Systems in Documentation, J. H.* Shera, A. Kent, and J. W. Perry, editors (Interscience Publishers, New York, 1957).

Hoyle, W. G., "Marginal Punched Cards for a Reference File in the Field of Electronics," *The Engineering Journal,* June 1958, pp. 61-66.

Breger, I. A., "Design of Simple Punched Card Systems with Reference to Geochemical Problems," *Economic Geology,* Vol. 53, No. 3, pp. 325–338 (May 1958).

Krieger, K. A., "A Punched-Card System for Chemical Literature," *Journal of Chemical Education,* Vol. 26, No. 3, pp. 163–166 (March 1949).

Tapia, E. W., "Coding Several Types of Chemical Patent Abstracts for Punch-Card Use," pp. 41–55 in *Progress Report in Chemical Literature Retrieval,* G. L. Peakes et al., editors (Interscience Publishers, New York, 1957).

Fleischer, M., "Experiences with a Notched-Card File of Geochemical Data," pp. 105-111 in *Progress Report in Chemical Literature Retrieval, G.* L. Peakes et al. (Interscience Publishers, New York, 1957).

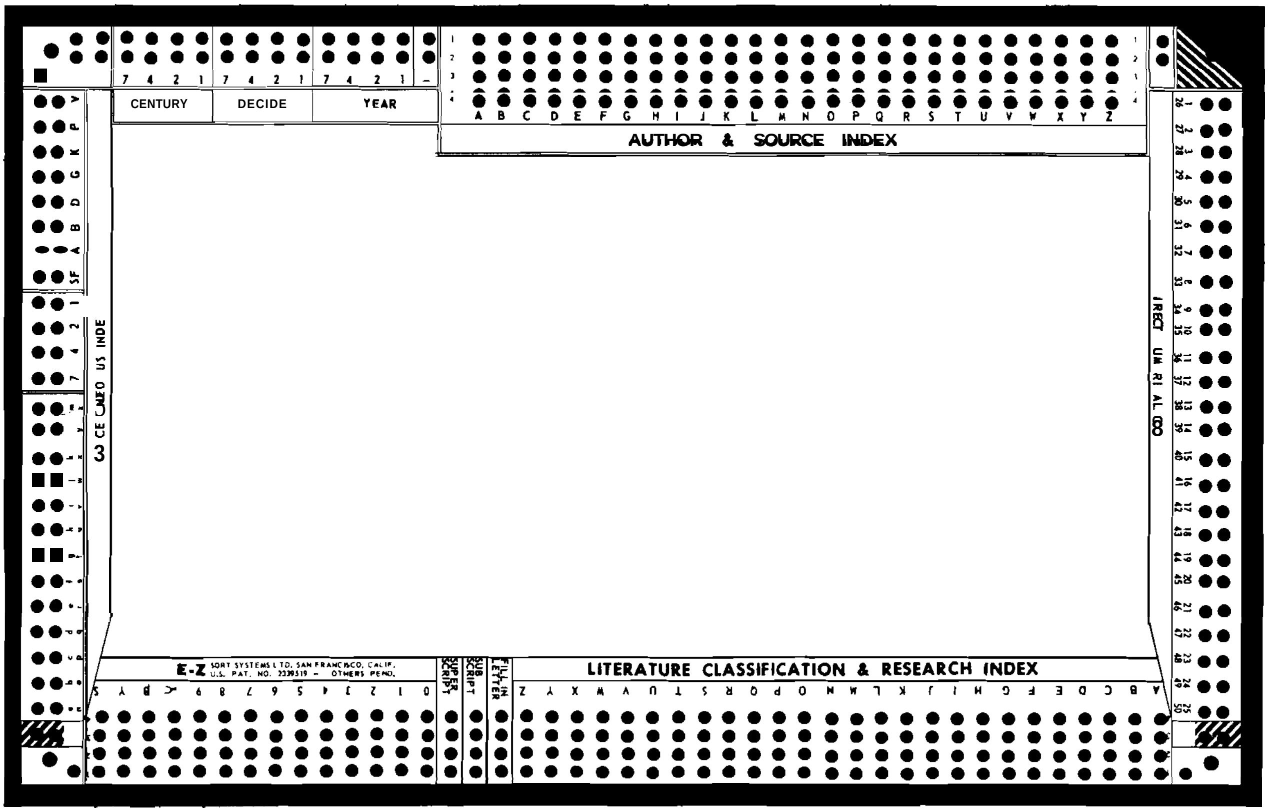


FIG. 5-35 Bibliography card.

meric punching of journal identification numbers, the direct alphabetic punching of a standardized journal title abbreviation, or by category punching, using a separate hole for each particular journal (if the number of journals of interest is not too large). The reference or abstract may be hand-written, typed, printed, duplicated, or pasted onto the face of the card. Figure 5-35 shows a general-purpose bibliographic card. This card has several numeric and alphanumeric fields that can be adapted by the user to implement his particular classification scheme.

Business Records. There are many ways in which notched cards can be used in the paper work system of a business organization. In addition to the examples already cited, notched cards are useful for the following purposes:

- General accounting and cost accounting
- Payroll distribution and check reconcilement
- Sales analysis
- Labor relations and union grievances
- General filing systems (e.g., correspondence records)
- Purchase orders and incoming orders
- Mailing lists
- Inventory control [15]
- Registration of warranties [16]
- Motor-pool trip reports
- Preventative maintenance records

Two specific examples of these applications are given below.

GENERAL ACCOUNTING. For a manual bookkeeping system with a large number of ledgers, it may be convenient to use a card for each double entry.

[15] McPherson, L. A., "Punch Cards for Inventory Control," *Hospital Management,* Vol. 83, No. 6, p. 86 (June 1957).

De Mambro, J., "Inventory Control for Electronic Parts Distributors," *National Electronics Distributors Association Journal,* July 1959, pp. 43–51.

[16] Nicholson, A. K., ('Automation of Warranty Cards Provides Fast Consumer Profiles," *Sales Management,* Vol. 84, No. 4, p. 48 (February 19, 1960).

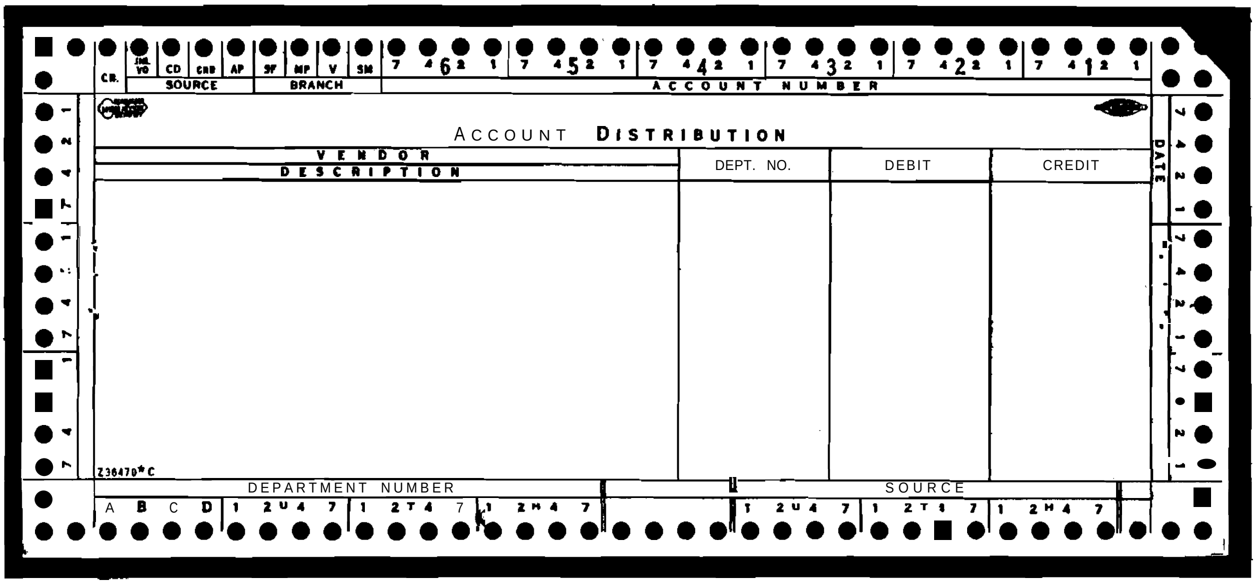


FIG. 5-36 General accounting card.

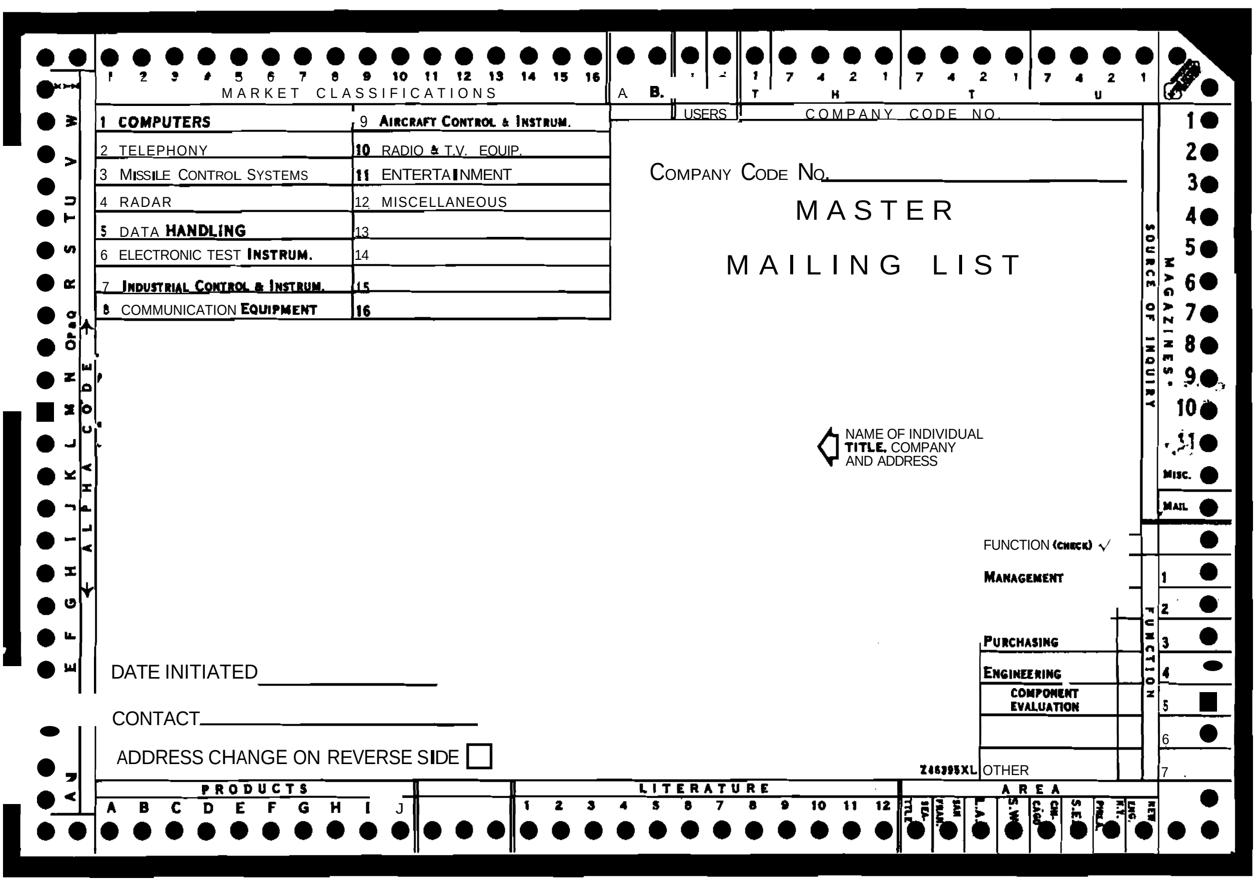


FIG. 5-37 Mailing list card.

Ledger or account numbers could be pre-punched so that when a batch of cards was received by the accounting department they could easily be sorted out into groups to simplify the posting to the ledgers. Such an application might also use the tabulating punch, described earlier in this chapter. An example of a general accounting card for a large department store is given in Fig. 5-36.

MAILING LISTS. Most companies maintain some sort of mailing list for direct mailing of promotional material, often organized by individual names, or company, or geographic location, or product interests. Cards can be used to advantage if different portions of the file are to be used for each mailing. For example, organizations that have a diverse product line and know the special interests of the people on their mailing list, would want to restrict their promotional mailings to those individuals who have an interest in the product to be described in a particular mailing. Figure 5-37 shows an example of a mailing list card maintained by an electronics component manufacturer.

Representative Costs

Notched cards may be ten times as expensive as IBM or Remington Rand cards, and two to three times as expensive as good catalog card stock, but the equipment for notching and manipulating the cards is relatively inexpensive. As mentioned earlier, notched card stock may even be prepared

TABLE 5-2

Representative Costs of Notched Card Equipment

Item	Approximate Rental Cost	Approximate Purchase Price
Electrofile (maaterkeyboard — 1-tray unit)	$ —	$1600
Electrofile (master keyboard—5-tray unit)	—	3000
Keymrt Tabulating Punch, Model 380	85/mo.	3410
Keymrt Hand Punch, Model 6201	—	8
Keymrt Data Punch. Model 8500	65/year	—
Keymrt Duplicating Punch	35/year	—
Keymrt Keypunch (manual), Model 8145	10/mo.	—
Keymrt Keypunch (electric), Model 8148	20/mo.	—
Keymrt Batch Groover (desk model 8203)	20/year	—
Keymrt Dupli Punch, Model 8300	48/year	—
Keymrt Groover. Model 8250	18/mo.	—
E-Z Sort Hand Groover	—	15
E Z Sorting Needles	—	1
Keysort Sorting Needle	—	2–4
E-Z Sorting Tray	—	8–13
Zator 800 Selector	45/mo.	—

TABLE 5-3

Representative Costs of Notched Cards and Card Sets

Item	Approximate Cost
3$1o-by-754-inch notched cards or stencil cards (for orders of 10,000 oards or less)	1 14 per thousand
3¾o-by-7½-inch notched cards or stencil carda (for orders of 100,000 carda)	6 per thousand
&by-&inch cards (for orders of 10,000 cards or less)	17 per thousand
5-by-8-inch cards (for orders of 100,000 carda)	8 per thousand
Zatocards	16 per thousand
Cards in coupon book form	100 to 200 per thousand books
Two-piece (3$1o-by-7 54-inch! sets with carbon-backed paper (for orders of 15,000 sets)	18 to 24 per thousand sets
Four-piece (3¾o-by-7½-inch) sets with carbon-backed paper (for orders of 15,000 sets)	28 to 32 per thousand sets
Card Savers	3 for 10 books
Practa Data-Cards	38 to 50 per thousand
Ekctrofik carda (5-by-8-inch)	60 to 68 per thousand

Note: Them are approximate prices. The exact price will depend upon the order quantity, card size, special printing requirements, and other features.

with tabulating card equipment by punching holes around the edges of a conventional IBM or Remington Rand card (see Fig. 5-25). An initial charge of 50 to 100 dollars is necessary to do the artwork and prepare the printing plate if a special card format is to be printed. Representative equipment costs are given in Table 5-2 and material costs in Table 5-3.

INTERIOR-NOTCHED CARDS

The rest of this chapter discusses cards (other than the IBM and Remington Rand type of tabulating card) that are punched in the interior of the card, and are primarily manual systems with a few machine aids. There seem to be two general types of interior-notched cards: those that are searched with needles in much the same manner as the edge-notched cards and those that are searched visually or optically (the so-called "peek-a-boo" systems).

Needled Cards

The needled cards are similar to the edge-notched cards in that the type of coding used for edge-notching is also used for interior-notching. The pre-punched holes are enlarged with hand punches or stronger desk-top punches. Searching is per-

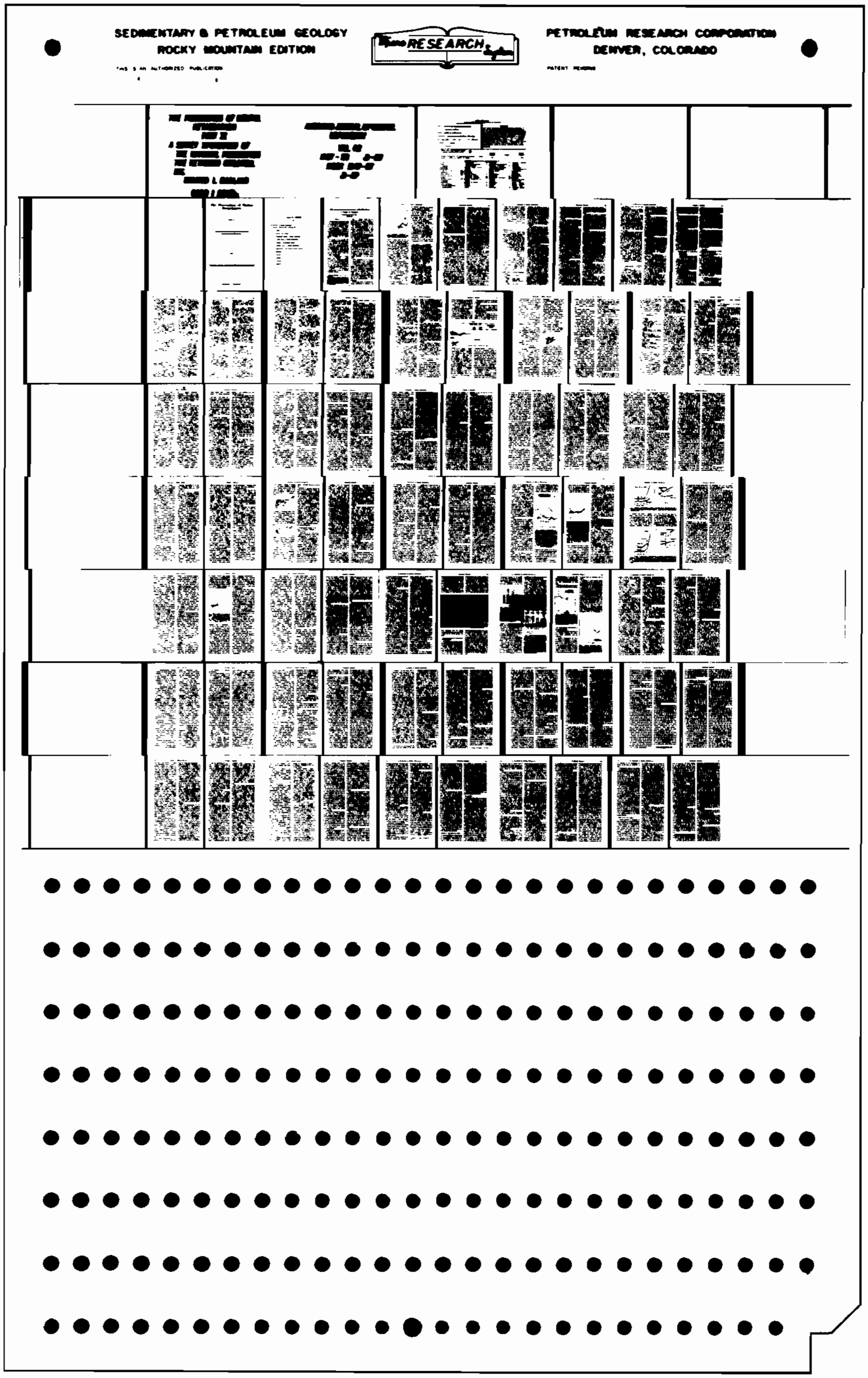


FIG. 5-38 Needled microfilm sheet of Petroleum Research Corporation.

formed by threading a long needle through the file, or fraction thereof, for each of the notched positions to be needled. By then removing the bottom support of the card deck, or by turning the entire deck upside down, the notched cards will be able to drop lower than the un-notched cards, and can then be extracted from the deck. An example of such a card is shown in Fig. 5-38.

The Ahrend-Globe interior-notched card system, with its relatively simple equipment, has been used rather extensively in Europe for many different types of applications.[17] One rather unusual application has been made in this country of an interior-needled card by the Petroleum Research Corporation.[18] Instead of using regular paper cards, they have applied this method to a file of sheet microfilm cards that they prepare and distribute on a subscription basis to persons interested in sedimentary and petroleum geology (see Fig. 5-38). Each article or report of potential interest to the service's subscribers is microfilmed and each page transferred to a 5-by-8-inch microfilm sheet, with up to 56 pages of text reproduced at about 1/18 its original size. One of these microfilm sheets is made for each subscriber. In addition, the indexing information for this article is punched into the microfiche so that each subscriber, with the use of a standard indexing system, obtains and maintains a complete reference file and search facility with very little effort on his part. Simple searching fixtures permit the subscriber to perform the file searching rather rapidly and easily.

Optical (Peek-a-Boo) Cards

Whereas the needled cards are often encoded in a large variety of ways, the optical coincidence cards are usually coded in a coordinate indexing manner. If coordinate indexing is used, then each item in the file is given a unique serial number or accession number, preferably starting with 1 and numbering consecutively. Each of the attributes, subject headings, or terms used for the indexing of the file contents is assigned a term card (see Fig. 5-39). Thus there is a term card for each descriptor, Uniterm, or subject heading which might be used to describe the items of information in the file. Each of these term cards contains a fixed number and pattern of assigned but unpunched hole positions on the interior of the card. Each of these hole positions is numbered according to a standard pattern, such as a rectangular checkerboard pattern. Each one of the hole positions represents a serial number, starting with 1 and counting consecutively up to the maximum number of holes that can be punched into a single card. If a particular term applied to a document, a hole would be punched in that term's card in the position corresponding to that document number. Consequently, a term card would be punched only with the numbers of the documents described by that term. A search for logical products of several terms (where the desired document is one that is indexed by term A and term B *and* term C) can easily be performed by superimposing the A, B, and C term cards and looking for the individual hole positions which are punched through all three cards. If the same hole position is punched on all three cards, it will be possible to see through the entire deck in that hole position. Thus, if a light is positioned at the rear of the deck, a pinpoint of light will show up in each hole position that has been punched for all three terms. In this way, the coordinates of the light spots indicate the serial numbers of the file items that satisfy the search request. This business of looking through the holes to see the light is the basis for the descriptive label of "peek-a-boo" systems. The basic ideas for this approach were developed in 1915 by Taylor, with some later modifications suggested by Soper.[19] An excellent description of the peek-a-boo approach and its application have been given previously by Wildhack and Stern.[20] For applications in which the file material and index are constantly changing, new cards may be prepared by a computer program.[21]

[17] Anon., "A New Approach to the ASM-SLA International Punched Card System," *American Documentation,* Vol. 10, No. 1, pp. 95–97 (January 1959).

[18] Chronic, J., "How Microfilm Library Aids Research," *World Oil,* Vol. 142, No. 6, pp. 95–97 (May 1956).

[19] Taylor, H., "Selective Device," U.S. Patent No. 1,165,-465, issued December 28, 1915.

Soper, H. E., "Means for Compiling Tabular and Statistical Data," U.S. Patent No. 1,351,692, issued August 31, 1920.

[20] Wildhack, W. A., and J. Stern, "The Peek-a-Boo System--Optical Coincidence Subject Cards in Information Searching,', Chapt. 6 in *Punched Cards,* 2nd ed., R. S. Casey et al., editors (Reinhold Publishing Corp., New York, 1958).

[21] Robbins, D. K., "Computer Production of Peek-a-Boo Sheets," *Communications of the Association for Computing Machinery,* Vol. 4, No. 12, pp. 562–565 (December 1961).

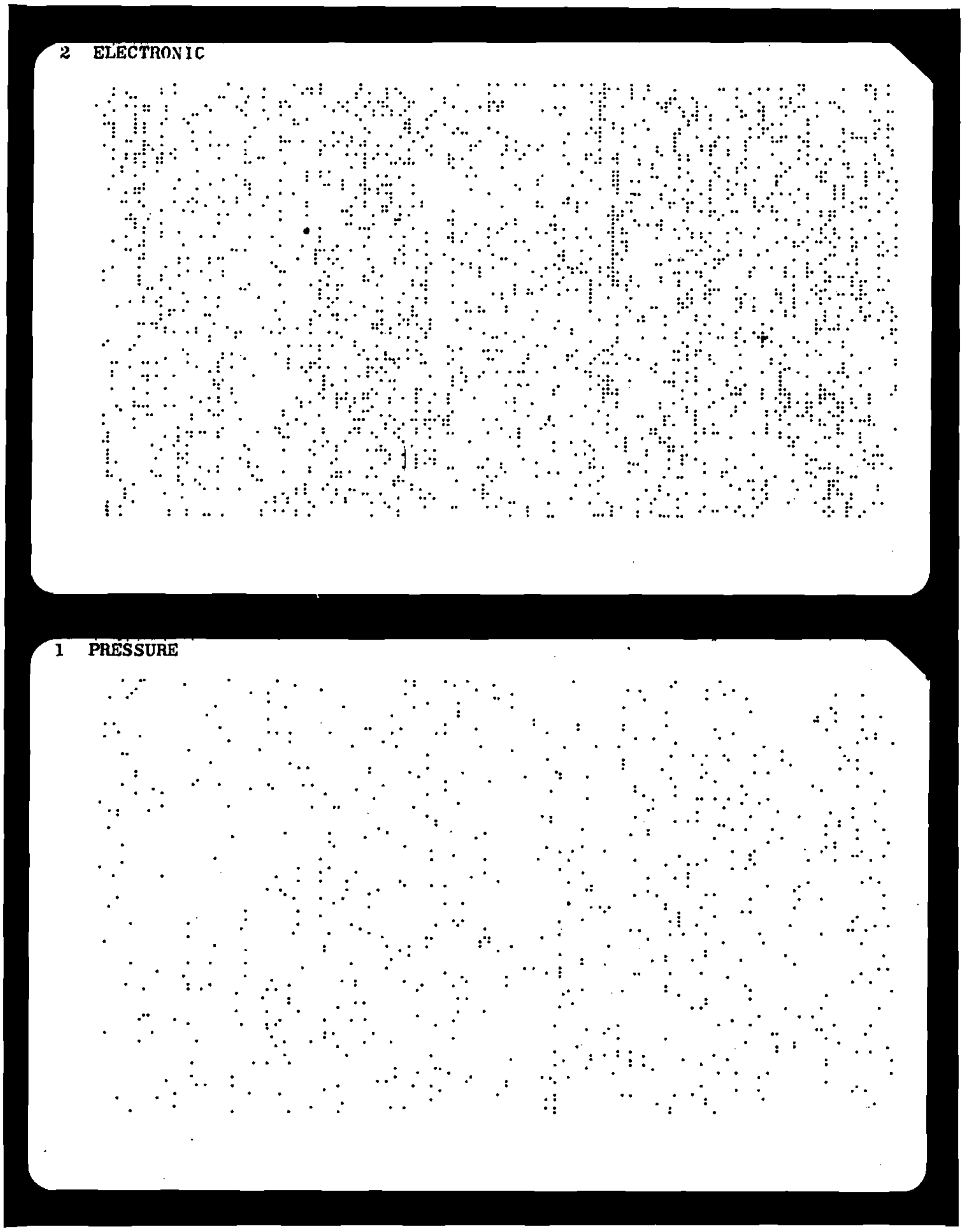


FIG. 5-39 Peak-a-boo term cards (National Bureau of Stondards).

The peek-a-boo cards may also be used as a data processing device for the evaluation of test data and the examination of relationships between many variables. It can easily provide cross-correlations and statistical analyses of survey or test results.[22]

[22] Campbell, J. D., and H. S. Caron, "Data Processing by Optical Coincidence," Science, Vol. 133, No. 3461, pp. 1333–1338 (April 28, 1961).

The use of pre-addressed hole positions on the card provides some restrictions because of the limited number of holes that can be punched into and read from a card. The file capacity can naturally be increased by increasing the card size and decreasing the size and spacing of the holes. The following card sizes have been used:

Number of Holes per Card	Size of Coding Field (in inches)	System
400	standard tab card size	Find-It
480	" " " "	IBM Port-A-Punch
540	" " " "	Omnidex
540	" " " "	Remington Rand card
960	" " " "	regular IBM card
10,000	9 x 9	Tennatrex
18.000	4 x 7¼	National Bureau of Standards
100,000	Microfilm filmstrip	Minimatrex

For some applications with very large files it is possible to divide the file into a number of separate sub-files, so that smaller peek-a-boo indexes can be operated in a parallel and independent manner. The Omnidex * system, described later, uses some additional encoding to allow large files of information to be represented by sub-files of 500-item cards. Several of these peek-a-boo systems are described in more detail below.

National Bureau of Standards Microcite System. One of the earliest applications of peek-a-boo cards to documentation problems was made by the U.S. National Bureau of Standards to index a collection of literature on instrumentation.[23] This initial application was followed by additional tests and the development of operating systems by other organizations.[24] The NBS term cards are 5-by-8-inch white plastic cards (see Fig. 5-39), and the equipment for punching and viewing is relatively simple (see Fig. 5-40). In most peek-a-boo systems the searcher usually has to view the light spots, record the numbers of the selected documents, and then obtain a copy of the document or abstract from permanent storage before he can actually appraise the search results. With the NBS Microcite feature, a microtransparency of the abstract of each document is placed on a microfilm array in the position corresponding to that document number. Thus, with the microfilm array placed behind the term cards during searching, any selected coordinate will now produce a microdisplay of a document's abstract instead of merely a dot of light. This microfilm image can be read with suitable magnifying lenses. A special searching machine has been built using this principle so that a user may more carefully examine the content of a selected document as a part of the search operation (see Fig. 5-41). With photocopy techniques, it is also possible to make a paper print of the display onto suitable paper so that the search results can be quickly recorded for later reference or mailing.

FIG. 5-40 National Bureau of Standards peek-a-boo punch and viewer in operation.

Termatrex. The first commercial versions of the National Bureau of Standards peek-a-boo system seem to have been Howard Benson's Omnidex equipment, put out in 1956, and the Termatrex equipment of Jonker Business Machines, Inc., who installed their first system in 1957.[25] The Termatrex equipment is a complete line of peek-a-boo equipment covering a wide range of file-size capabilities and input-output devices. The simplest system is the Model 104. For input, the cards corresponding

* Trademark.

23 Wildhack and Stern, chapter in Casey et al., cited earlier in this section, fn. 20.

Wildhack, W. A., and J. Stern, "The Peek-a-Boo System in the Field of Instrumentation," Chapt. 13 in Information Systems in Documentation, J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

Wildhack, W. A., J. Stern, and J. Smith, "Documentation in Instrumentation," American Documentation, Vol. 5, No. 4, pp. 223–237 (October 1954).

24 Wright, R. C., and C. W. J. Wilson, "Classification with Peek-a-Boo for Indexing Documents on Aerodynamics: An Experiment in Retrieval," Proceedings of the International Conference on Scientific Information, Vol. 1, pp. 771–801 (National Academy of Sciences, Washington, D. C., 1959).

25 H. F. Benson, "File It to Find It . . . Omnidex," an instruction manual of Omnidex, Pomona, Calif., 1959.

Jonker, F., "The Termatrex Inverted 'Punched Card' Systems," American Documentation, Vol. 11, No. 4, pp. 305–315 (October 1960).

FIG. 5-41 National Bureau of Standards peek-a-boo image viewer.

to the index terms of that document are placed together on a drilling template, and a hole is drilled with a portable hand drill through all the cards simultaneously at the hole number corresponding to the serial number or accession number of the document. To simplify this drilling process, two modified versions of this machine, Model 202 and Model 301 (Figs. 5-42 and 5-43), provide a drill-holding mechanism which is suspended above the cards, travels on two perpendicular tracks, and can be locked into position during the drilling operation to ensure accuracy in the hole drilling. The numbering system for the cards in this machine is a simple rectangular coordinate system of 100 by 100 points, resulting in a total of 10,000 hole positions on a card. For searching, the cards corresponding to the search terms are superimposed on a viewing screen and the base light illuminates the punched addresses that are common to all the cards. The coordinates of the holes can be read by means of movable scales. A similar system (Keydex) by the Royal McBee Corporation was introduced in 1963.

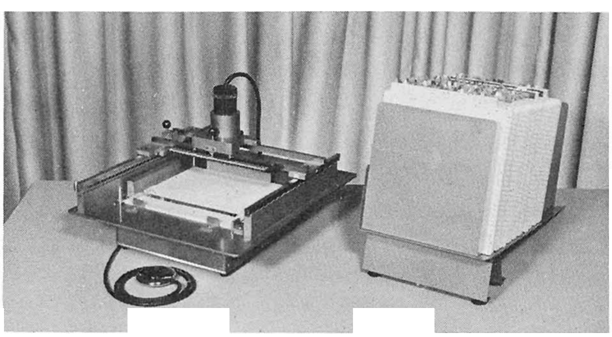


FIG. 5-42 Jonker Model 202 retrieval equipment.

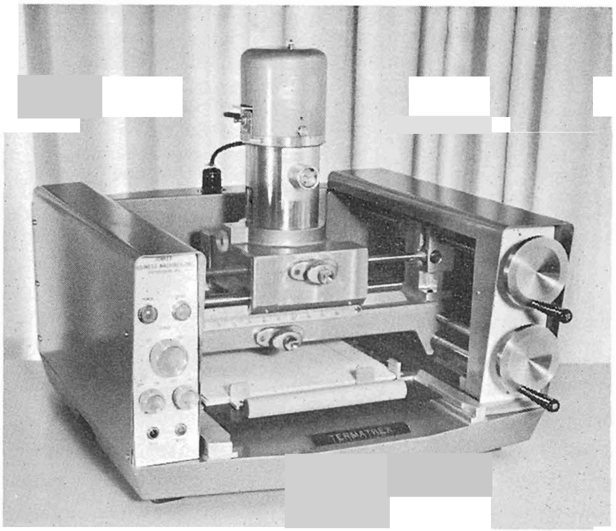


FIG. 5-43 Jonker Model 301 retrieval equipment.

A more complex system, the Jonker Model 400 (see Fig. 5-44), is available for the applications with unusually heavy input and search traffic, and utilizes auxiliary input equipment. A keyboard entry device is available which causes a card to be punched in the position corresponding to the coordinates entered into the keyboard. Input could also be provided from tab cards or punched paper tape. To simplify the recording of search results, equipment is also available which will print or punch out the hole number when a cross-hair sighting arrangement is placed over that hole, or when an automatic scanning device locates an illuminated hole. Equipment is also available, Jonker Model

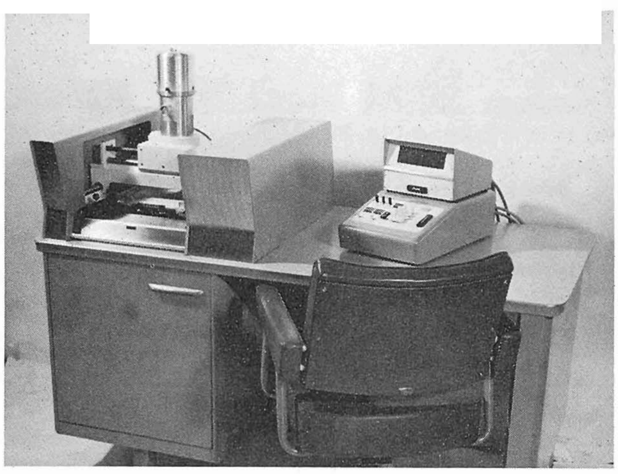


FIG. 5-44 Jonker Model 400 retrieval equipment.

81, to obtain a permanent photographic record of the results of the superimposition of a deck of cards.

For applications that require a larger file capacity, Jonker has developed a microfilm peek-a-boo system, the Minimatrex, in which reduced images of standard 10,000 item cards are stored on microfilm strips ("termstrips"). Each filmstrip contains 5 or 10 separate frames, with each frame a photographic reduction of a standard Termatrex card. An early model of the special viewer is shown in Fig. 5-45, and a display of one of the frames of a termstrip is shown in Fig. 5-46. Up to 12 termstrips can be superimposed for a single search on this viewer. The termstrips would probably have to be made with special microfilm equipment or techniques, in order to achieve the necessary control of distortion and registration.

Several users have indicated that about 3 minutes of machine time is required to enter a document into the system, and that the search time for a file of 10,000 or 40,000 items is usually on the order of 2 to 4 minutes. This compares very favorably with computer searching speeds, and is probably less expensive. In addition, any of the peek-a-boo systems have the advantage of being available for almost immediate searching at any time during the day—whereas, searches on a computer system are usually batched and run once per day.[26]

FIG. 5-45 Jonker Model 1000 visual Minirnatrex reader.

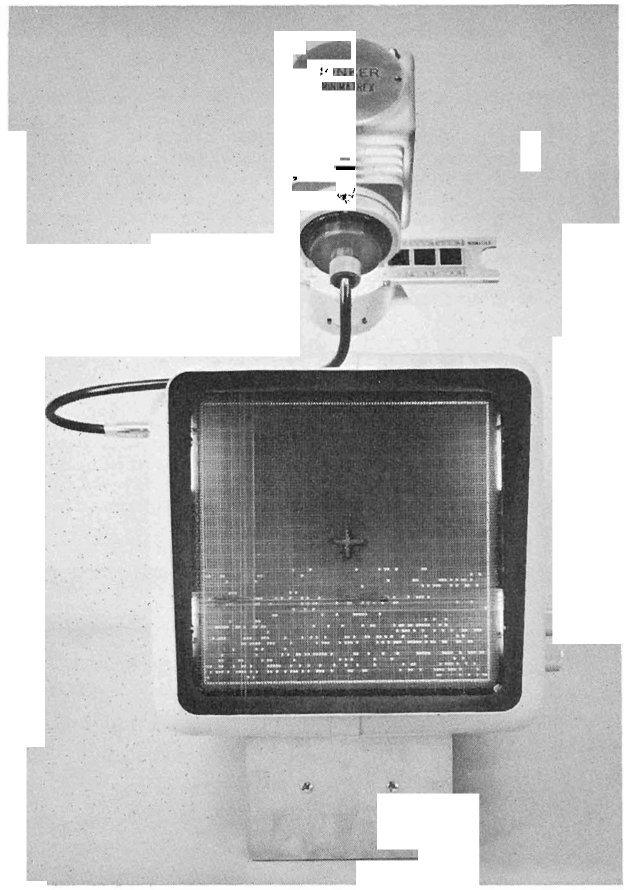


FIG. 5-46 Jonker Model 1000 viewing screen.

Omnidex. The Omnidex peek-a-boo system uses a card stock with the dimensions of a standard tabulating punched card, and an array of 12 rows and 45 columns of pilot holes pre-punched into the card (see Fig. 5-47). This pattern provides 540 hole positions on the card, so that if each position were assigned to a particular file item, the file would be limited to a collection of 540 items. However, the card format is arranged so that 40 of the holes are used to denote a system of sub-files. In this manner, a master deck restricts the search ta bne or more of 40 sub-files (each with a capacity of 500 items). The selected sub-files are subsequently examined to

[26] Carpenter, H. M., "A System for Storage and Retrieval of Data from Autopsies," ***American Journal of Clinical Pathology,*** Vol. **38,** No. 5, pp. 449–467 (November **1962).**

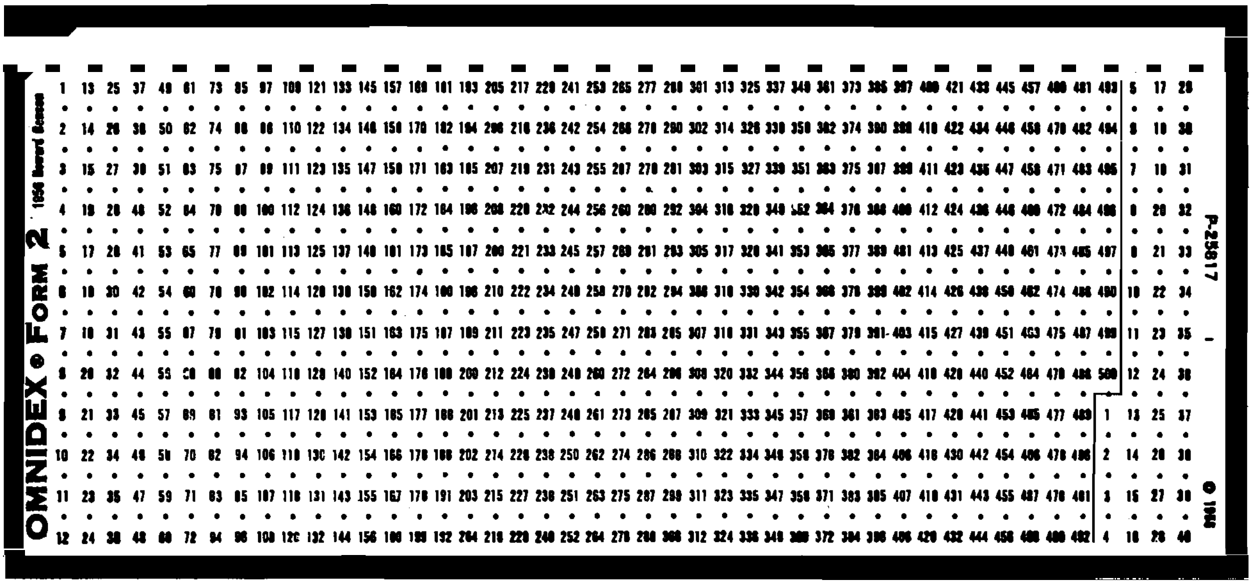


FIG. 5-47 Omnidex card.

locate the exact reference number. This scheme permits the system to index up to 20,000 items. For still larger collections, combinational coding (2 holes out of 40) can be used to control up to 250,000 items in 500 file groups of 500 documents each. The pre-punched pilot holes are used to guide the hand punching and assure the proper alignment of punched holes. This is possible because the hand punch is constructed in such a manner as to prevent punching unless a pilot needle is centered in the pilot hole. Although the Omnidex system has received relatively little attention to date, its first commercial application (17,500 items) occurred in 1956 at the General Dynamics–Pomona Engineering Document Center. Installations have been made in about 25 major companies since that time.

IBM *Port-A-Punch.* The IBM Port-A-Punch card described in an earlier chapter can be used conveniently for a small peek-a-boo system. Each card has 480 pre-scored hole positions, and a hole can be put in a card simply by pushing it out with a pencil point. However, relatively little use has been made, so far, of these small peek-a-boo systems. In some restricted applications, a regular IBM card has been used.[27] A commercial version

[27] Malinoski, R., et al., ***Use** of **the Cardatype Machine to Prepare Peek-a-Boo Cards,*** Report ***30*** of the Groth Institute, Pennsylvania State University, University Park, Pennsylvania (August 7, 19591, ASTIA Report No. 228 899.

TABLE 5-4

Representative Costs of Interior-Punched Card Equipment

Item	Approximate Purchase Price
Jonker Model 100	$ 390
Jonker Model 202	1,350
Jonker Model 301	4,000
Jonker Model 400	12,000
Jonker Model 500 Automatic Card Scanner	10,000
Jonker Model 1000 Visual Minimatrex Reader	1,500–2,000
Jonker Model 81 Photological Device	450
Omnidex Hand Punch	30
Royal McBee Keydex	1,200

TABLE 5-5

Representative Costs of Interior-Punched Cards

Item	Approximate Cost
Omnidex cards	$25 per thousand
IBM Port–A-Punch cards	3 per thousand
Find-It Card Kit	10 per two hundred
Jonker Termatrex cards (plain index tabs)	25 per hundred
Jonker Termatrex cards (random access index)	35 per hundred

TABLE 5-6

Commercially Available Interior-Punched Card Systems

Card or System Name	Distributor
Termatrex	Jonker Business Machines, Inc. P.O. Box 265 404 North Frederick Avenue Gaithersburg, Maryland
Trio	Ahrend-Globe Simon-Stevinweg 7–21 Hilversum, Netherlands
Omnidex	Howard Benson P.O. Box 825 Pomona, California
Findex	William K. Walthers, Inc. 1245 N. Water Street Milwaukee 2, Wisconsin
Brisch-Vistem Feature Card	Visirecord, Inc. (Distributors for Carter-Parratt, LM., London) 375 Park Avenue New York 22, New York
Keydex	Royal McBee Corp. 850 Third Avenue New York 22, New York
Find-It	Find-It P.O. Box 36074 Wilshire-La Brea Station Los Angeles 34, California

of this system is marketed by Find-It (see Tables 5-5, 5-6) providing 400 hole positions.

Brisch-Vistem Feature Card. The Brisch-Vistem Feature Card system is another peek-a-boo system for relatively small collections.[28] A card measuring 6 by 11 inches provides 1000 hole positions, while a card that measures 10 by 12 inches contains *10,000* hole positions. Very simple equipment is used for punching, and no special equipment is provided for searching or viewing.

[28] Jolley, J. L., "Punched Feature Cards," ***Data Processing*** (British), April-June 1959, pp. 4–11.
Johnson, A., "Experience in the Use of Unit Concept Coordinate Indexing Applied to Technical Reports," ***Journal of Documentation,*** Vol. 15, No. 3, pp. 146–155 (September 1959).

Representative Costs

The interior-notched cards are generally more expensive than regular punched tabulating cards. However, the total equipment investment for interior-notched cards is usually very modest. Representative equipment costs are given in Table 5-4, and material costs in Table 5-5. A list of the distributors or manufacturers of this equipment is given in Table 5-6.

Additional References

Aronoff, S., "Code for Alphabetical Index for Punched Cards," ***Journal of Chemical Education,*** Vol. 36, No. 11, p. 581 (November 1959).

Ball, N. T., "Making Classification Systems for Punched-Card Coding," Chapt. 25 in ***Punched Cards,*** 2nd ed., Casey et al., editors, cited below.

Casey, R. S., C. F. Bailey, and G. J. Cox, "Punch Card Techniques and Applications," ***Journal of Chemical Education,*** Vol. 23, No. 10, pp. 495–499 (October 1946).

Casey, R. S., J. W. Perry, A. Kent, and M. Berry, editors, ***Punched Cards: Their Application to Science and Industry,*** 2nd ed. (Reinhold Publishing Corp., New York, 1958).

Cox, G. J., R. S. Casey, and C. F. Bailey, "Recent Developments in Keysort Cards," ***Journal of Chemical Education,*** Vol. 24, No. 2, pp. 65–70 (February 1947).

Hoffer, J. R., "Information Retrieval in Social Welfare: Experiences with an Edge-Notched Information Retrieval System," ***American Documentation,*** Vol. 13, No. 2, pp. 169–175 (April 1962).

McGaw, H. F., ***Marginal Punched Cards in College and Research Libraries*** (The Scarecrow Press, Washington, D.C., 1952).

Royal McBee Corp., ***Keysort Notching and Sorting,*** Brochure S-604 [Port Chester, New York (undated)].

Scheele, M., ***Punch-Card Methods in Research and Documentation: With Special Reference to Biology, J. E.*** Holmstrom, translator (Interscience Publishers, a division of John Wiley and Sons, New York, 1962).

Thompson, M. S., "Peek-a-Boo Index for a Broad-Subject Collection," ***American Documentation,*** Vol. 13, No. 2, pp. 187–196 (April 1962).

The Equipment
Representative Applications
Representative Costs

6

Punched card **systems**

THE EQUIPMENT

This chapter deals with the IBM and Remington Rand punched tabulating card (tab card) equipment, and the ways in which it can be used for information storage and retrieval applications. Since the actual card form and methods of coding were described earlier, this chapter will concentrate primarily on functional descriptions of appropriate card equipment, indications of what costs can be expected, and descriptions of many applications of this equipment to storage and retrieval problems. Although most of the equipment illustrated happens to be IBM equipment, corresponding equipment is available from Remington Rand in nearly all cases. Remington Rand and IBM each offers some equipment for which the other has no equivalent. In most instances, the figures in this chapter illustrate one particular model of each machine type under discussion.

Key Punch

The most common method of creating tab cards is to use a key punch unit similar to that shown in Fig. 6-1. Simpler manual card punches, such as the models shown in Figs. 6-2 and 6-3, are also available but operate at much slower speeds. The operator of the key punch depresses the keys on a device similar to a typewriter keyboard, and the machine automatically punches the corresponding holes into a card. The card is advanced automatically after punching each column, and new cards are fed in from an input hopper when called for. The key punches may also be equipped with printing mechanisms that can print the punched information at the top of the card, above the corresponding columns. In this way, the entire contents of the card can be printed in one line at the top of the card at the same time that the card is being punched. If necessary, the key punch can automatically enter any

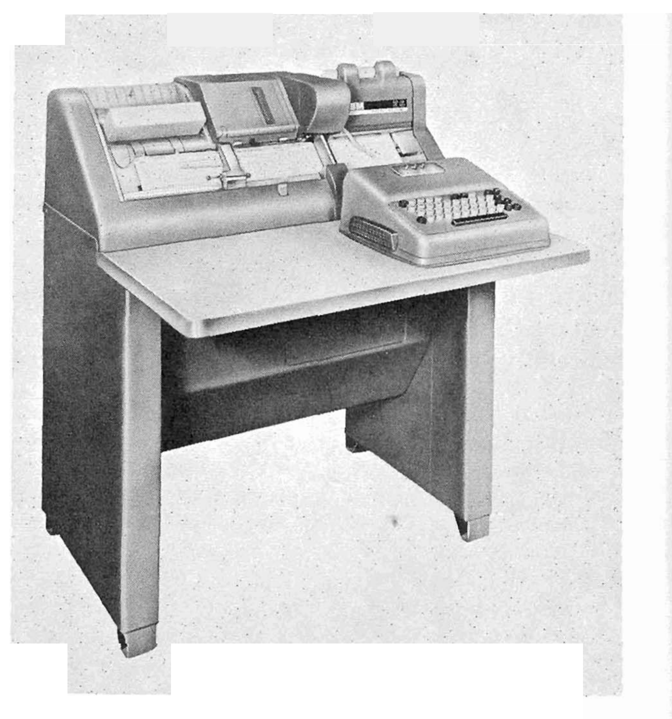


FIG. 6-1 Card punch (IBM Model 026).

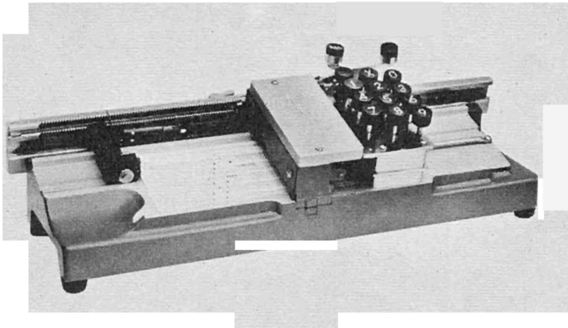


FIG. 6-2 Manual key punch (Panels, Wires, Inc., Model 001).

amount of fixed or constant data into each card as it is being punched, or can rapidly skip sections of the card that are to be left blank. The operator thus needs only to key in the variable data. Any or all of the data that were punched into a card can also be duplicated automatically into the next following card. Skilled key punch operators can punch the cards as fast as 10,000 numeric characters per hour (125 full cards per hour) if the conditions are right. A more reasonable estimate for personnel or time requirements is 5000 numeric or 2500 alphanumeric characters per hour, using a fairly well-trained operator.

It is almost certain that there will be some mistakes in the punching, just as a typist can always be expected to make an occasional mistake. Consequently, another card machine, the verifier, is used to detect these mistakes. An average of about 2 percent of punched cards contain errors detected by the verifier, while about 2 cards in 10,000 will subsequently be found to contain errors not detected by the verifier.[1]

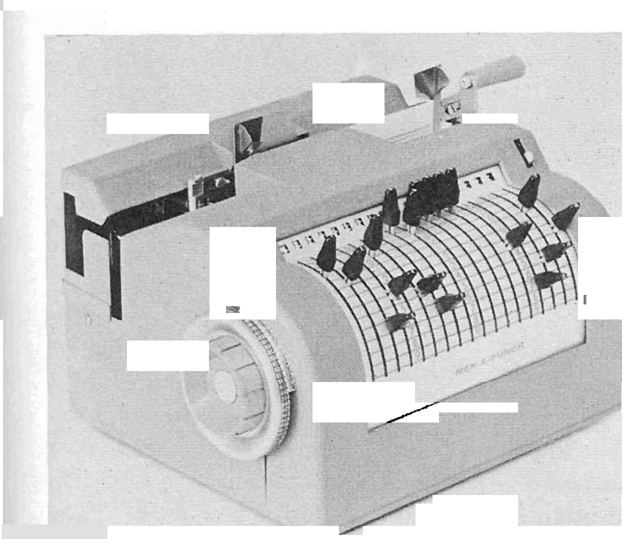


FIG. 6-3 Manual key punch (American Electronics, Inc., Mek-A-Punch).

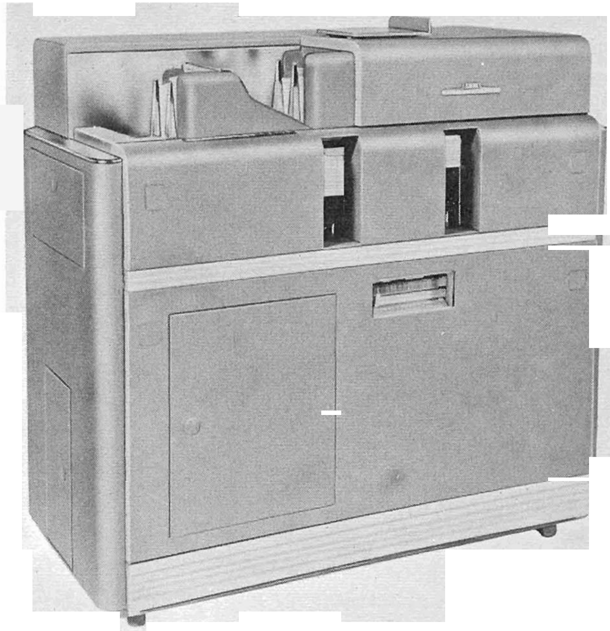


FIG. 6-4 Reproducing punch (IBM Model 514).

Verifier

The verifier machine looks much like the key punch machine. The punched cards to be verified are placed in the input hopper, and the operator proceeds to use the keyboard and the original source data in the same manner as the original key punch operator. As a further check for errors, the cards are usually verified by a different operator from the one who originally punched them. For each column, the verifier compares the character of the depressed key with the character already punched in the card. A difference between these two will cause the machine to stop at that point so that the operator can observe the discrepancy. The operator can then pass the card through the madhine, but the top of the card will be notched above the column in doubt. Whenever a card successfully passes a check

[1] Hinds, G., "Accuracy of Data Preparation," *Computer* Bulletin (British), June 1960, pp. 7–9.

by the verifier, it automatically receives a notch on one edge of the card. This provides a means for later determining whether or not the card has been verified. Most organizations that use key punches also use verifiers to check the card punching.

Reproducing Punch

The reproducing punch (see Fig. 6-4) provides a means by which cards can easily be duplicated, or fixed information can be punched into all of the cards in a deck. One deck of previously punched cards is fed into one hopper of the reproducing punch while a deck of blank cards is fed into another hopper. A blank card is then punched to conform exactly to each card, or specified portion thereof, in the original deck. Decks of cards that are wearing out can be regenerated, or additional copies of cards can be generated for subsequent filing, distribution, or parallel processing. When reproducing a deck, the machine automatically checks the corresponding punched cards against each other, and indicates the discrepancies when they occur. A plugboard wiring arrangement permits the user to program the machine to transcribe any or all of the data in the original card to different columns of the new card, or to duplicate only a selected portion of the original card. A plugboard or patchboard (see Fig. 6-5) is a control panel similar in nature to a telephone switchboard and is used to permit changes to be made relatively easily in the internal electrical circuits of the machine in order to change the sequence of operations of the equipment. The actual physical wiring or interconnection of the electrical circuits of the equipment needs to be changed for each different computation or operation to be performed. The plugboard offers a convenient way to plug in jumper wires to make these necessary connections. The entire plugboard and its patchwork of wires can be easily removed from the equipment and replaced with another plugboard. Thus for jobs that are run frequently, a separate plugboard may be maintained and stored, with all its wiring intact.

Most tab card equipment incorporates some form of plugboard, since it provides a relatively inexpensive way to change the sequence of operations of a piece of equipment. However, a considerable amount of time and training are usually necessary before a user can properly prepare a plugboard. In addition, plugboards are limited in the number of sequence steps that they can control, and they are bulky and difficult to store. Consequently, as the data processing equipment gets more complex (e.g., computers) the sequence of machine operations for each job may be controlled by computer programs stored in the machine, rather than by plugboards. This is discussed in more detail in the next chapter.

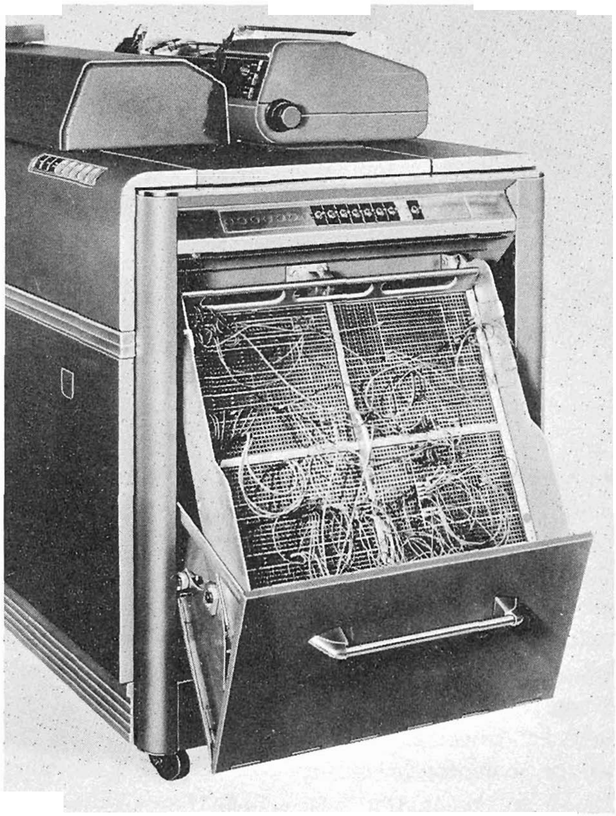


FIG. 6-5 Representative plugboard.

The reproducing punch can interpret the penciled mark-sense information on a card and punch it into the same card. One model of the reproducing punch can print pre-wired or pre-punched numbers across one end of each punched card. These printed numbers are in large, bold type, which is appropriate for, say, employee time cards, where individual cards must be recognized quickly. Another model of the reproducing punch has an arithmetic capability for accumulating totals from information punched into the cards, and punching the total into the original data card or into a new card. Some of these reproducing punches have been adopted for use as input-output devices for computer systems. This type of equipment normally punches at the rate of 100 or 200 cards per minute, although some units operate at 300 cards per minute.

Inferpreter

For many applications, it may be convenient to have printed on the top of the card the same information that has been punched into that card. In some cases, the information can be printed at the top of the card when it is keypunched. However, the printing of a key punch unit (see Fig. 4-2) is not of the best quality, and can only appear at the top of the card. In addition, there is no convenient way to suppress the printing of specified columns on the card, and the printing will not be available if the decks are generated by some means (e.g., reproduced with a reproducing punch) other than keypunching. The interpreter is a machine for printing numeric or alphanumeric data from selected portions of a card onto selected positions of the same card. Because of its improved print quality (see Fig. 6-17), this equipment will permit tab cards to be used as documents, notices, or entries in a manual reference file. The interpreter, shown in Fig. 6-6, operates at a rate of 60 cards per minute.

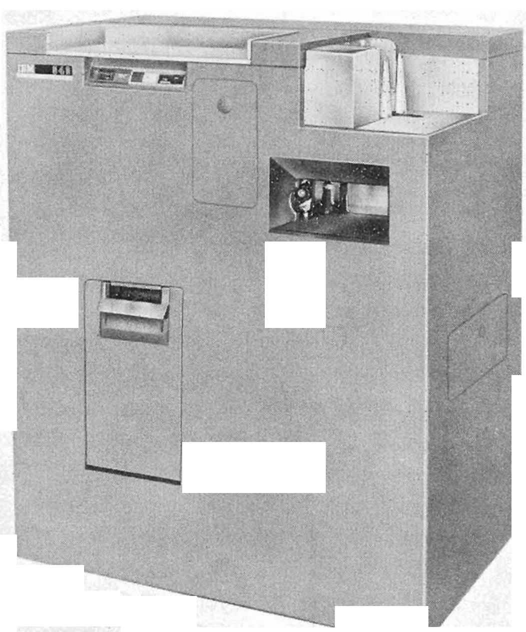


FIG. 6-6 Alphabetic interpreter (IBM Model 548).

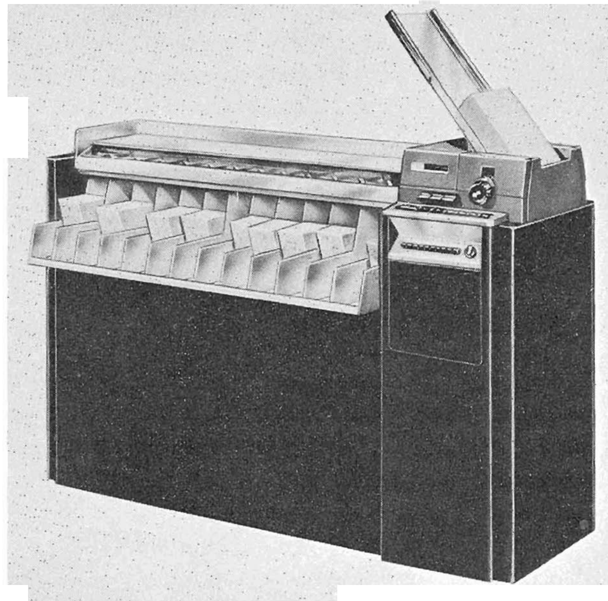


FIG. 6-7 Sorter (IBM Model 084).

Sorter

The sorter provides a means by which groups of cards may be sorted into numeric or alphabetic sequence. It also provides a means by which cards can be selected from a file on the basis of the information which is contained in the card. Several sorter models are available; one is shown in Fig. 6-7. Most sorters examine only one specified column position on the card while the entire group of cards is passed sequentially through the machine, and direct each card to the hopper corresponding to the character punched on the card. The machine operator designates the column to be examined by setting selector switches or buttons on the machine. Because a deck of cards can normally be sorted by only one column at a time, the deck must be passed through the sorter as many times as there are characters in the sorting or selection field.[2] A deck would have to be passed through the sorter ten times in order to be arranged in sequence by a ten-digit number. The number of passes is doubled for alphanumeric data. Some sorters link mechanical counters to each of the sort pockets or hoppers so that the cards in each sorted

[2] The IBM Model 101 Statistical Sorter, described in more detail in a later section, is the major exception to this restriction to single-column sorting, and consequently provides many additional capabilities.

category can be tallied. Various single-column sorter models are available that will operate at rates of 450, 650, 1000, and 2000 cards per minute.

Collator

The collator is a machine with a moderate amount of logical capability for card-handling operations. It can select specific cards from a deck, merge two decks of cards into a single ordered deck, and match one deck against another. In a single pass through the machine, it can select cards by the following criteria:

Cards punched with specific digits
Cards with a blank column
Cards with a specific number
Cards higher than a specific number
Cards lower than a specific number
Cards between two specific numbers
Cards that are out of sequence

The collator's merging operation, in one machine pass, combines two sets of ordered cards into one set of cards in an ordered sequence. Both sets must originally be arranged in the same sequence (e.g., both in ascending order) before they are merged. The collator is often used for file-maintenance operations to insert a small amount of information into an existing file. New data can usually be incorporated into an existing file much faster by a merging than by re-sorting the entire group of new and old cards. The collator's matching operation is used to compare one card file to another to check, for example, that a purported duplicate deck is actually duplicate; or to extract cards from a master file which corresponds to a set of interrogation cards. Some of the newer collators operate at speeds up to 1300 cards per minute. One collator is shown in Fig. 6-8.

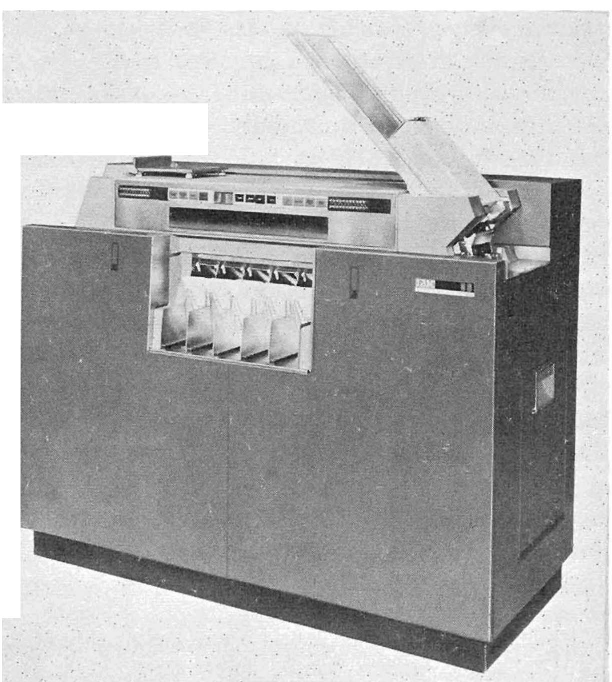


FIG. *6-8* Collator (IBM Model *088).*

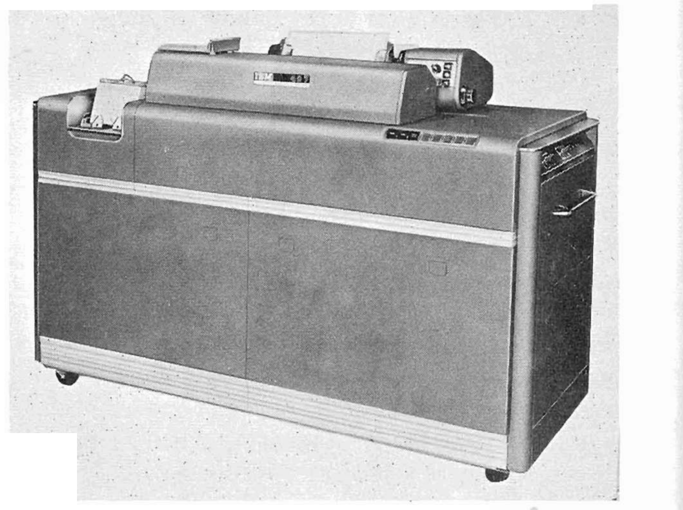


FIG. *6-9* Accounting machine (IBM Model *407).*

Line Printer (or Accounting Machine)

The line printer (see Fig. 6-9) derives its name from the fact that it prints one complete line at a time, instead of printing each of the line's characters sequentially. An illustration of the type font and quality of such printing is given in Fig. 6-16. The first machines developed to print a line at a time from tab card information were designed as accounting machines with a capability to read information from cards, perform some computations, and print the results, along with other information from the card or from plugboard wiring, in complete lines. For many information processing tasks, little or no use is made of this available arithmetic capability, and the machine is used entirely for "listing," or printing the information that is punched into a deck of tab cards onto a sheet of paper. A plugboard is used with the machine so that it can be wired to read from designated portions of the card and position the designated information at various locations across the line of print. Usually, the information from one card is distributed along one line of print, although other arrangements can be devised. One machine (IBM Model 409) performs the dual function of punching data in the first 24 columns of a card at the same time that it prints some information on the remainder of the card. This is a very convenient way to generate documents for distribution (e.g., checks, utility bills, motor vehicle registration) which are to be read by humans and then returned (in whole or in part) for additional machine processing. A punched paper tape control unit on the machine provides some control over the paper movement, and can initiate such functions as the insertion of a predetermined number of line spaces at appropriate times. The internal logic of the machine can also be used for, say, automatically counting and printing page numbers or page headings. Line printers from IBM are normally available with 120 characters per line; each character position can draw from the symbols described in Fig. 4-2, and special characters can be substituted if desired. One modification of a computer line printer (IBM Model 1403) provides a full alphabet of both uppercase and lowercase letters. Most tab card printers normally work at a speed of 150 lines per minute, regardless of the amount of information to be printed. However, the line printers used with computer equipment often print 600 or 900 lines per minute, and sometimes more.

Tape-fo-Card Punch

This machine (see Fig. 6-10) provides a means by which the information on a punched paper tape can be transferred to cards. Similar equipment is also available to transfer information from cards to tape. Most equipment will read only the particular paper tape code for which they were designed, although a few models of more complex equipment can read almost any paper tape and transcribe the information to tab cards. The paper tape may have been generated by devices such as teletype tape punches, or it may have been made on tape-punching typewriters. The paper tape often contains additional information to control the actions of the machine, in addition to the data to be transferred to the card. Once the machine has started to read the tape, it can continue to operate automatically under control of the tape. With an optional keyboard added, this machine may also be operated as a regular key punch. The equipment usually operates at a maximum speed of about 20 card columns per second, and represents one way to conveniently bridge the gap between manual operations and machine systems. More specific information about paper tape equipment is given in Chapt. 8.

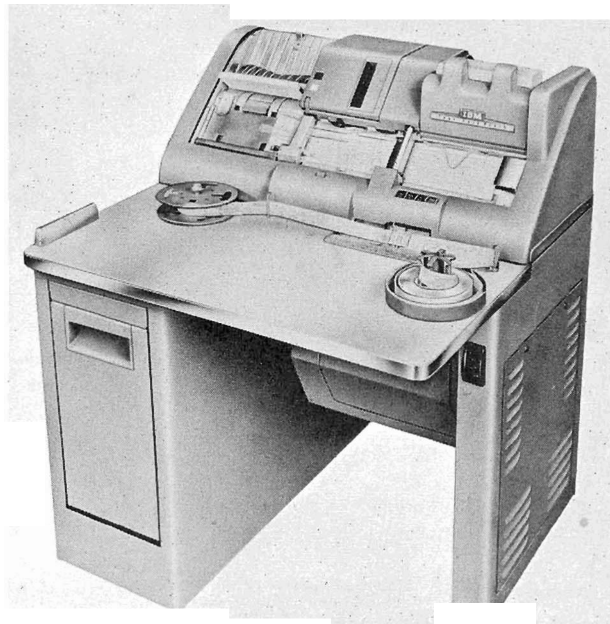


FIG. 6-10 Tape-to-card punch (IBM Model 046).

Statistical Sorter

This versatile machine (see Fig. 6-11) has the capability to sort, count, accumulate, edit, and print summaries from tab card data. The sorter can arrange the cards in any desired pattern, and can search files of cards by specific search criteria, or by a combination of specific criteria. The statistical sorter can sort the cards one column at a time, as can all other sorters. However, plugboard wiring enables this machine to perform more complex sort operations than the other sorters. The flexibility provided by the plugboard is somewhat offset by the added inconvenience, since a plugboard wiring layout must be made for every special search operation on the statistical sorter. This machine is espe-

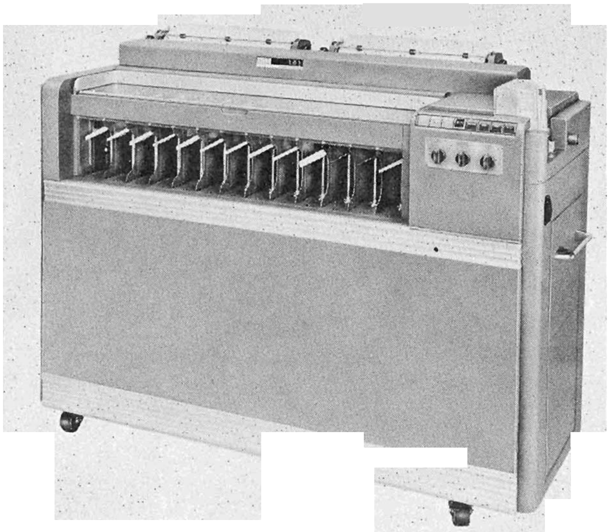


FIG. 6-11 Statistical sorter (IBM Model 101).

cially efficient for file searching operations.[3] In one pass of the deck it can select all cards that have the same punching as a master card. This machine is also useful for searching by ranges of variables. For example, if a continuous range of temperatures were punched into the cards, the cards could be separated into groups or ranges of temperatures (e.g., 0°–20°, 21°–40°, 41°–80°, 81°–500°, and everything over 500°) by one machine pass. The machine can also be used for checking the sequence of a deck of cards, sorting a deck of cards by variable-length numbers or words, selecting the nth card of a group, and selecting all unpunched cards. The statistical sorter operates at a rate of 450 cards per minute.

[3] Garfield, E., "Preliminary Report on the Mechanical Analysis of Information by Use of the **101** Statistical Punched-Card Machine," *American Documentation,* Vol. **5,** No. **1,** pp. **7–12** (January **1954).**

Secrest, B. W., *The IBM* 101 *Electronic Statistical Machine Applied to Word Analysis of the Dead Sea Scrolls,* an un-numbered report of the IBM World Trade Corp., New York (November **17, 1958).**

Jones, W. S., and P. H. Butterfield, "A Technical Information Service Using Punched Cards for Indexing and Retrieval," in *Progress Report in Chemical Literature Retrieval, G.* L. Peakes et al., editors (Interscience Publishers, New York, **1957).**

Schultz, C. **K.,** "An Application of Random Codes for Literature Searching," Chapt. **10** in *Punched Cards,* 2nd ed., R. S. Casey et al., editors (Reinhold Publishing Corp., New York, **1958).**

Maierson, A. T., and W. W. Howell, "Application of Standard Business Machine Punched-Card Equipment to Metallurgical Literature Reference," *American Documentation,* Vol. **4,** No. 1, pp. **3–13** (January **1953).**

COMAC Mark-2

The COMAC Mark-2 (see Fig. 6-12), developed by Benson-Lehner Corp. in 1962, is an improved commercial version of the Continuous Multiple Access Collator (COMAC) developed by Documentation, Inc., in 1957.[4] An earlier commercial version, the Special Index Analyzer (IBM Model 9900) was developed by IBM Corp. in 1958, but less than six

[4] Taube, M., "The COMAC: An Efficient Punched Card Collating System for the Storage and Retrieval of Information," Chapt. **5** in Vol. **5** of *Coordinate Indexing* (compiled by *M.* Taube, Documentation, Inc., Washington, D.C., **1959).** Also published in the *Proceedings of the International Conference on Scientific Information,* Vol. **2,** pp. **1246–1254** (National Academy of Sciences, Washington, D.C., **1959).**

Taube, M., "Experiments with the **IBM-9900 and** a Discussion of an Improved COMAC as Suggested by These Experiments," *Journal of Chemical Documentation,* Vol. **2,** No. 1, pp. **22–26** (January **1962).**

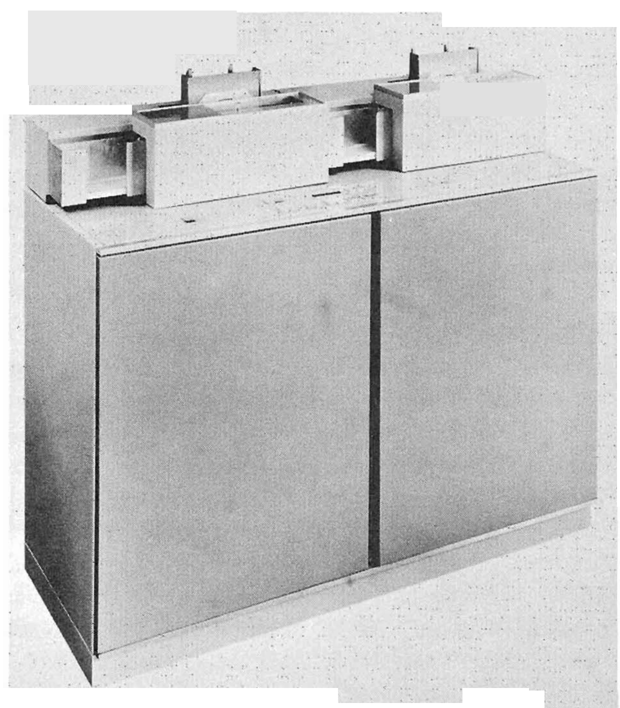


FIG. 6-12 COMAC Mark-2.

units were installed.[5] All of this equipment was designed primarily for searching files of punched cards, and was especially intended for collections of data or technical documents that had been indexed by a coordinate indexing system. The systems use inverted coding, with a set of one or more tab cards collected together to list the document numbers of all reports indexed by a particular term. The term decks are updated by adding the new document numbers into each of the term sets that describe those particular documents. Searching is performed by manually extracting and then machine-collating the sets of the search terms together, two at a time, to obtain the document numbers which satisfy the search logic. This is the same approach described earlier in connection with the collator. The machines can find logical sums (i.e., "term A OR term B"), logical products (i.e., term A AND term B), complements (i.e., NOT term A), and some combinations of these operations. The result of the search is a list of selected document numbers. On the IBM 9900 this list can be printed out on an electric typewriter. On the COMAC Mark-2 this list can be automatically punched out into tab cards by a key punch unit. The entire file collection can normally be searched in less time than a serial scanning of the file would take because only a few cards (the pertinent term sets) are actually passed through the equipment. In addition, each card of a term set can hold up to 12 six-digit document numbers. The COMAC Mark-2 has some additional features, such as the ability to duplicate portions of its file if necessary, and the ability to read a group of term cards that are punched with one document number per term card and condense this information into new terms cards that are punched with up to 12 numbers each.

Universal Card Scanner

The card scanner (Fig. 6-13) was developed for file-searching operations.[6] This machine examines

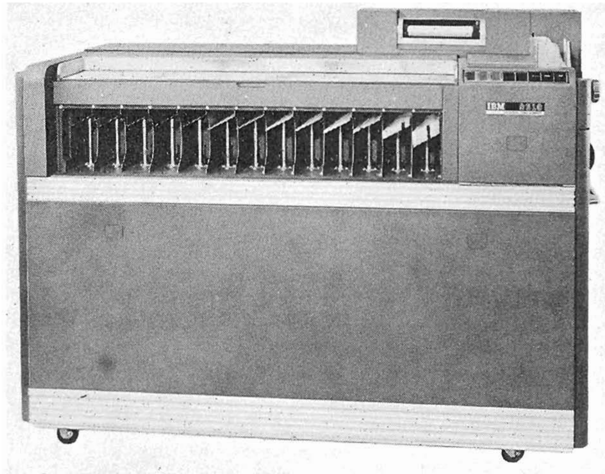


FIG. 6-13 Universal card scanner (IBM Model 9310).

cards one at a time, and notices whether any of one or more search patterns are wholly or partially contained in any of the cards which are scanned. The machine uses a punched question-card to furnish the search patterns to be used for comparison with the deck of cards to be searched. File cards that match all or part of the search pattern are directed into separate pockets on the sorter. With this type of selector or scanner, the entire card file must usually be passed through the machine, rather than selected portions of the file as is done with the COMAC. The system is therefore described as using direct instead of inverted coding. The statistical sorter can also be equipped with a similar scanning attachment.? Less than six of these special units have been installed.

Interrelated Logic Accumulating Scanner (ILAS)

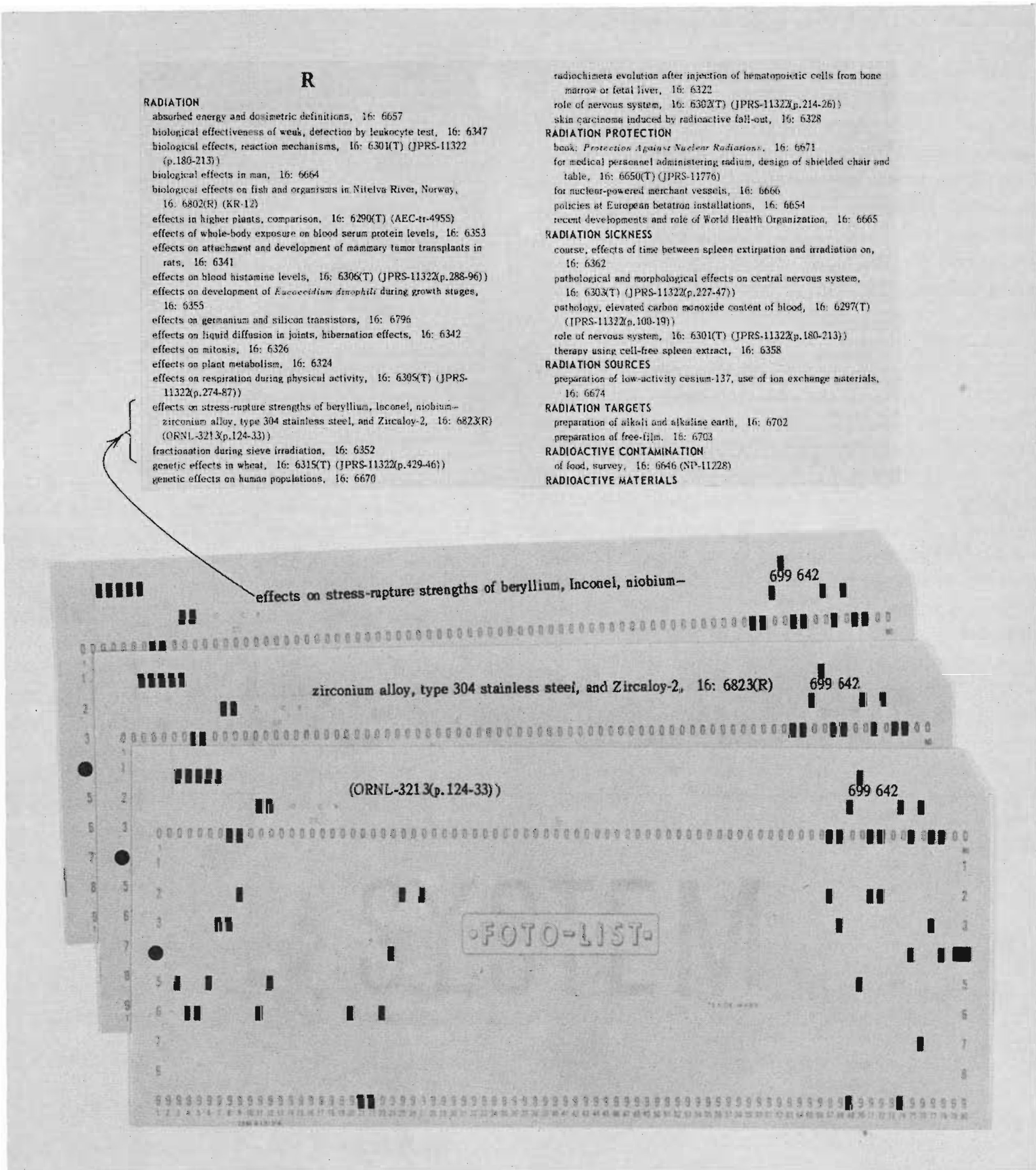
ILAS, a special punched card scanner, was designed by the Office of Research and Development

[5] Murphy, R. W., *The ZBM 9900 Special Index Analyzer,* a non-numbered report from the IBM Corp., Poughkeepsie, New York (November *17, 1958).* Also printed as Chapt. 6 in *Coordinate Indexing,* Vol. *5,* compiled by M. Taube (Documentation, Inc., Washington, D.C., *1959).*

[6] Luhn, H. P., *The ZBM Universal Card Scanner for Punched Card Znformation Searching Systems,* a non-numbered report from the IBM Research Center, Yorktown Heights, New York (November *17, 1958).* Also printed as Chapt. *7* in *Coordinate Indexing,* Vol. *5,* compiled by M. Taube (Documentation, Inc., Washington, D.C., *1959).*

[7] Luhn, H. P., *Row-by-Row Scanning Systems for ZBM Punched Cards as Applied to Znformation Retrieval Problems,* Report RC-100 from the IBM Research Center, Yorktown Heights, New York (May *1959).*

Luhn, H. P., *Information Retrieval Through Row-by-Row Scanning on the IBM 101 Electronic Statistical Machine (Row-by-Row Scanning Aattachment),* an un-numbered report from the IBM Research Center, Yorktown Heights, New York (November *17, 1958).*



R

RADIATION
absorbed energy and dosimetric definitions, 16: 6657
biological effectiveness of weak, detection by leukocyte test, 16: 6347
biological effects, reaction mechanisms, 16: 6301(T) (JPRS-11322 (p.180-213))
biological effects in man, 16: 6664
biological effects on fish and organisms in Nitelva River, Norway, 16: 6802(R) (KR-12)
effects in higher plants, comparison, 16: 6290(T) (AEC-tr-4955)
effects of whole-body exposure on blood serum protein levels, 16: 6353
effects on attachment and development of mammary tumor transplants in rats, 16: 6341
effects on blood histamine levels, 16: 6306(T) (JPRS-11322(p.288-96))
effects on development of *Eucoccidium dinophili* during growth stages, 16: 6355
effects on germanium and silicon transistors, 16: 6796
effects on liquid diffusion in joints, hibernation effects, 16: 6342
effects on mitosis, 16: 6326
effects on plant metabolism, 16: 6324
effects on respiration during physical activity, 16: 6305(T) (JPRS-11322(p.274-87))
effects on stress-rupture strengths of beryllium, Inconel, niobium–zirconium alloy, type 304 stainless steel, and Zircaloy-2, 16: 6823(R) (ORNL-3213(p.124-33))
fractionation during sieve irradiation, 16: 6352
genetic effects in wheat, 16: 6315(T) (JPRS-11322(p.429-46))
genetic effects on human populations, 16: 6670
radiochimera evolution after injection of hematopoietic cells from bone marrow or fetal liver, 16: 6322
role of nervous system, 16: 6302(T) (JPRS-11322(p.214-26))
skin carcinoma induced by radioactive fall-out, 16: 6328

RADIATION PROTECTION
book: *Protection Against Nuclear Radiations*, 16: 6671
for medical personnel administering radium, design of shielded chair and table, 16: 6650(T) (JPRS-11776)
for nuclear-powered merchant vessels, 16: 6666
policies at European betatron installations, 16: 6654
recent developments and role of World Health Organization, 16: 6665

RADIATION SICKNESS
course, effects of time between spleen extirpation and irradiation on, 16: 6362
pathological and morphological effects on central nervous system, 16: 6303(T) (JPRS-11322(p.227-47))
pathology, elevated carbon monoxide content of blood, 16: 6297(T) (JPRS-11322(p.100-19))
role of nervous system, 16: 6301(T) (JPRS-11322(p.180-213))
therapy using cell-free spleen extract, 16: 6358

RADIATION SOURCES
preparation of low-activity cesium-137, use of ion exchange materials, 16: 6674

RADIATION TARGETS
preparation of alkali and alkaline earth, 16: 6702
preparation of free-film, 16: 6703

RADIOACTIVE CONTAMINATION
of food, survey, 16: 6646 (NP-11228)

RADIOACTIVE MATERIALS

FIG. 6-14 Fotolist cards ready for filming and a sample of the resulting copy.

of the U.S. Patent Office for experimental work in chemical patent searching. Only one unit has been built, and has been in operation at the U.S. Patent Office since about 1956.[8] It is mentioned here primarily because of its historical value.

Card-Actuated Cameras

Many indexes, directories, and catalogs are produced today by typing directly onto high-quality white punched cards (see Fig. 6-14), arranging the cards in proper sequence, and then running the cards through a special camera such as the one in Fig. 6-15 to take a high-speed picture of the material that is typed on the top of the card. The special cameras, such as VariTyper Fotolist, Recordak Listomatic, and Friden Compos-0-Line, sequentially expose a band across a strip of negative film which is one page column wide. The cameras usually photograph one typed line from the top of each card, although the Listomatic camera will automatically adjust its aperture according to the punches in the card, in order to expose one, two, or three lines from that particular card. In the example shown in Fig. 6-14, the information is typed one line per card, and then photographed after the cards have been assembled in proper order by tab equipment. The exposed film is then used to prepare an offset plate for printing. One attractive feature of printing from a card is the fact that the cards can be mechanically sorted and arranged into order before the filming is done. Also, after being photographed in one type of arrangement, the cards may then be sorted into some other arrangement and then re-photographed. Publication of updated lists may be accomplished by adding and deleting cards from the original files by manual or machine methods, and then re-photographing the new card file. The cards may also be used for other tab machine operations if necessary. Another attractive feature is the fact that the typing on the card may be in any desired font, and is not necessarily restricted to 80 characters as an IBM card is. In practice, any kind of graphical information may be placed on the top of the card for photographing, although as mentioned earlier, there is the restriction

[8] Andrews, D. D., et al., "Recent Advances in Patent Office Searching: Steroid Compounds and ILAS," Chapt. 25 in *Information Systems in Documentation*, J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

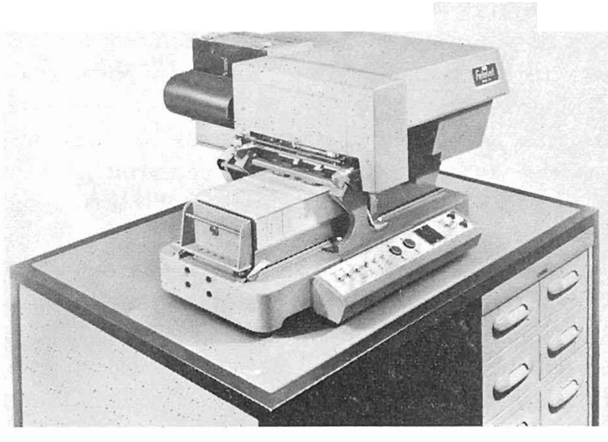


FIG. 6-15 Fotolist camera (Model 90).

that it be placed within certain locations on the card.

The cards are usually typed on electric typewriters or VariTyper machines which have been modified with special card-holding platens in order to ensure accurate positioning of the characters on the card. Special holes are provided in the margins of some of the cards to ensure proper alignment in the typewriter.

The Listomatic camera can operate at a maximum rate of 13,800 cards per hour, whereas the Compos-0-Line and Fotolist Model 970 can operate at approximately 7200 cards per hour. This technique of composing from tab cards is currently being used to produce the *Index Medicus*, the world's largest title announcement bulletin.[9] It is also being used to prepare the index to *Nuclear Science Abstracts* and *Chemical Abstracts*.[10] At least three other serial publications *(Electronics, Nucleonics, Textile World)* are using card-actuated cameras to prepare their directories or indexes.[11] The equipment is also being used for other commercial com-

[9] Miller, E. A., "The National Library of Medicine Index Mechanization Project," *Bulletin of the Medical Library Association*, Vol. 49, No. 1, Part 2 of 2 parts (January 1961).

[10] Day, M. S., and I. Lebow, "New Indexing Pattern for Nuclear Science Abstracts, *American Documentation*, Vol. *11*, No. 2, pp. 120–127 (April 1960).

Dyson, G. M., "Current Research at Chemical Abstracts," *Journal of Chemical Documentation*, Vol. *1*, No. 1, pp. 24–28 (January 1961).

[11] Markus, J., "State of the Art of Published Indexes," *American Documentation*, Vol. *13*, No. 1, pp. 15–30 (January 1962).

position work such as the Dun and Bradstreet business directories, telephone directories, parts lists, price lists, and maintenance manuals.[12] Some commercial printing shops have installed card-actuated cameras for job-shop production of catalogs, directories, and other material.[13] Some printers have even used the equipment to prepare entire training manuals and reports.

For simple copy and using these card-actuated cameras, the total cost to prepare a finished film copy from an original typed manuscript is about 6 to 20 cents per line, whereas the cost to prepare a duplicating master for the same copy using a tab listing is about 2 cents per line. If the user types his own cards he may then have them run through a camera at a service bureau at a cost of about 2 cents per line. A good typist can prepare 100 to 150 cards per hour.

REPRESENTATIVE APPLICATIONS

There are a number of ways in which tab equipment can be used to advantage in a library or other type of information system. The applications described or suggested to date, seem to fall into three general categories: (1) those in which composition and printing are the major objectives; (2) those in which file searching is the major objective; and (3) those other types of operations that are normally considered to be part of the processing or administrative job of the information system. Each of these three general areas is described in more detail in the following sections.

Situations Where Composition and Printing Are the Major Objectives

Tab card equipment has been used a great deal to prepare catalogs, directories, lists, and other tabulations. This is usually done by "listing" the cards on the line printer (see Fig. 6-9) after they have been arranged in the desired order. The line printer can print directly on paper forms with a few carbon copies for direct distribution, or on paper masters for spirit or offset printing of many copies. The arrangement of information on the page is determined by the layout of information on the cards, the sequence of cards, and the wiring of the plugboard on the printer. Consequently, there is a considerable amount of flexibility in the way in which the information can be arranged on the printed page. However, the tab equipment does have definite limitations such as the absence of lower case letters, the limited number of symbols available, the single type font and the difficulty of changing fonts in a machine, the possibility of occasional printing errors due to equipment malfunctions, and the aesthetically poor print quality. Where print quality is important, tab cards can be used with the card-actuated cameras described earlier. There may also be limitations to the use of tab equipment because of the restricted amount of data that can be stored in a single card, although additional cards (sometimes called "trailer" cards) could be used. However, it is more cumbersome and takes more machine time to handle sets of two or more cards per file item instead of a single card.

Card equipment is especially useful when a published listing is modified so often that a revised list must frequently be published and distributed. Company telephone books, personnel rosters, Uniterm dictionaries, classification schedules, and printed library catalogs are examples of this type of situation. A sample page of a company telephone directory which is prepared and maintained in this manner is shown in Fig. 6-16. Changes and additions to the list can be made simply by replacing and adding new cards to the file in their proper locations. Up-to-date lists can be obtained at any time, merely by listing the cards on the line printer.

Union Lists and Lists of Library Holdings. Lists of library holdings can be generated in a relatively straightforward manner by punching one or more identification cards for each file item, arranging the cards in some desired sequence, and then listing the cards on the printer. In the case of a list of journals or periodicals, a library might punch a card such as the one shown in Fig. 6-17 for each journal, indicating such things as the title, language, frequency of

[12] Anon., "D & B Shoots Its Directories," *Business* Automation, May 1962, pp. 40–41.

Anon., "Cameras, Cards Create Catalogs," Systems Management, April 1961, p. 26.

[13] Anon., "From Punched Card to Printed Page," Management and Business Automation, August 1960, pp. 28–29.

Prioletta, A., "New Photomechanical Equipment Speeds Preparation of Lists, Directories, and Text," OffsetDuplicator Review: Reproductions Reference Guide, 1st ed., pp. 208–211 (1960).

Minor, R., "Cards and Camera Speed Catalog Production," Management and Business Automation, April 1960, pp. 35–37.

EXT	ALT	NAME	SPOUSE	TITLE	LOC	ORG
2266	2796	FULLMFR, DALE 6069 PARK BLVD., P.A.	/	DATA REDUCTION AIDE DA 1-2682	320	654
2813	3342	FULTON, JOHN K. 340 MFADOWOOD DRIVE* M.P.	/SHIRLEY	SECT MANAGER FORMULA UL 1-2958	F238	463
2387	3530	FURBUSH, PETER 602 SPRING ST., R.C.	/ELAINE	BUYER EM 9-6062	G153	062
3159	3372	FUREY, DON L. 3941 CALIFORNIA AVE., APT. 19, M.V.	/	RES ENGINEER YO 8-5326	410B	646
2145	2141	FURGER, CASOL 380 IRENE CT., BELMONT	/	SECRETARY LY 1-4817	3102	254
2047	3140	FURIO, KIZO 967 967 ROBLE ST., APT.3, M.P.	/LOLA	CHMN DEPT SOLID STAT DA 6-0689	106A	455
3474	2664	FURIYO, ALBERT 3101 UNIVERSITY AVE., P.A.	/ELIZABETH	CHMN MEDICAL SCIENCE DA 3-8055	100B	539
2222	2223	FURMAN, GEORGE G. 3390 BAY RD., P.A.	/LOIS	GFN FOREMAN DA 2-2735	300	091
3475	2771	FURVERI, ELLEN I. 610 ADOBE PL., P.A.	/MARK	ASST FOR REPORT SERV DA 1-5929	G113	403
3255	2234	BURRER, MARK 533 E. 34TH ST., APT. 15-0, N.Y.	/J	INDUSTRIAL ECONOMIST OR 9-6564	N.Y.	203
2432	2373	FURST, FRANK 688 WESTRIDGE, P.P.	/LOUISE	PHYSICAL CHEMIST 32 2-1429	G224	452
2811	2159	FURTADO, MANUEL S. 773 CONNEMARA WAY, SUNNYVALE	/LORY	RES ENGINEER RE 6-1324	406B	644
2643		FURUSHO, YOSH 630 DEODARA DR., L.A.	/CHIYOKO	MACHINE SHOP FOREMAN YO 7-1574	300	405
3219		FURUYA, JOHN J. 6646 ARMOUR DR., SANTA CLARA	/MARGARET	SUPV MAIL SERVICES	E023	079
2222	2223	FUSS, CARL C. 719 WARRINGTON AVE., R.C.	/BETTY JEAN	LABORER EM 9-3167	300	091
2665		FUSSELL, RICHARD J. 993 CEREZA DR., P.A.	/LEE	DEVELOPMENT ENGR DA 3-5503	410B	632
3450	3449	FUTERNICK, JAMES 556 HEDGE RD., M.P.	/DIANE	MGR PUBLIC RELATIONS DA 3-5407	C113	006
3221	3220	FUTTRELLE, BARBARA 951 LAYNE CT., APT. 5, P.A.	/	KEYPUNCH OPERATOR 32 6-9486	G158	032
2077		FUTTERER, CHARLES A. 3784 EDGEWOOD RD., R.C.	/	ORGANIC CHEMIST EM 8-2402	102A	494
	3388	FUZY, JULIE 483 TYRELLA ST., M.V.	/DAN	PHYSIOLOGIST	100	539
3484	2838	FYFEE, WILLIAM A. 1061 PARTRIDGE* M.P.	/MARY	RES METEOROLOGIST DA 5-6058	404A	653
2475		FYLER, MARGARET 420 EMERSON ST., P.A.	/	REPORT TYPIST DA 2-3335	A019	403
3134	2437	GAAB, ROBERT M. MAPLE AVE., TAYLORSVILLE, KENTUCKY	/	SR ELECTRONICS TECHN GR 7-4953	308B	654
2570	2632	GABANY, ALEX 356 BOSTON AVE., APT. 4, S.J.	/BARBARA ANN	LAB TECHNICIAN CY 5-9570	104B	486
2642		GABEL, RAYMOND F. 5097 SEYFIRTH WAY, S.J.	/REY	SR MODEL MAKER AN 6-9079	300	405
2349	3164	GABLER, RICHARD W. 514 STANFORD AVE., P.A.	/HOPE	MECHANICAL TECHN DA 6-4548	414	646
2222	2223	GABOR, JULIE 3227 MADERA AVE., M.P.	/LENAL	MATRON DA 4-6534	300	092
2659	3164	GABRILLSON, HARLEY P. 861 LA YESA DR., M.P.	/LINDY	RES ENGINEER DA 3-5491	410A	646

FIG. 6-16 Example of a company telephone directory prepared by punched card equipment.

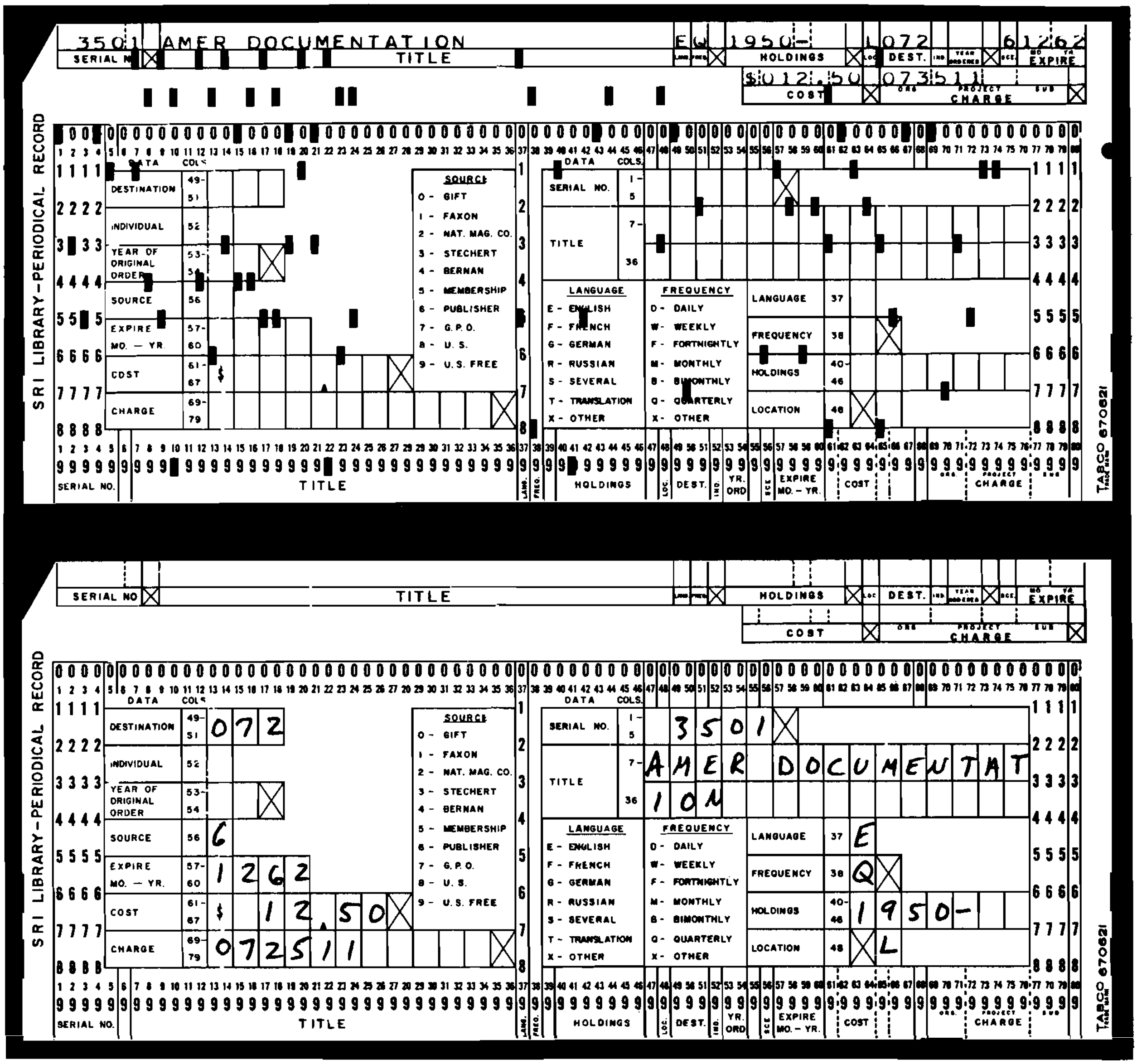


FIG. 6-17 Punched cards used for library lists of periodicals.

publication, earliest issue in the library, location of holdings, vender or source, expiration date, subscription fee, and cost accounting charge number. The cards could be sorted and then listed on the tabulator to obtain an alphabetized list of holdings. The same cards could be used to easily prepare special reports on such things as:

List of titles, their location, and extent of holdings—total or by sub-library
Subscriptions addressed for routing to particular offices
Subscriptions received by each office
Number of subscriptions
Amount spent by each department for subscriptions
Number of issues checked in during a given time period
Number of new subscriptions placed each year
List of new subscriptions expiring on a given date
Expected monthly billing for subscriptions
List of publications received from each vender or service.

In one such operating system, the tab card (see Fig. 6-17) has been especially designed for this purpose so that data for the key punch operator can be initially written on the card itself, and the cards can be kept up to date by making notes on the punched cards as the changes occur. In this way, the blank card is used as a convenient printed form for recording the data, and the punched card is also used as a form for manually recording corrections. The sample cards of Fig. 6-17 show the card used for the initial manual recording of data, and for the subsequent recording of the data. Some of the punched information is printed at the top of the card by an interpreter machine in order to facilitate manual filing. Several other libraries have utilized tab equipment for the control of library subscriptions.[14]

If the cards for two or more libraries can be punched in the same format, then a "Union List of Periodical Holdings" can be prepared simply by merging the cards from all the participating libraries, sorting into the desired sequence, and printing a common listing. This approach has been used to prepare union lists for several groups of cooperating libraries. It has also been used to prepare New Serial Titles, a continuing supplement to the Library of Congress's Union List of *Serials*.[15] The same approach could be extended to file material other than periodicals. The actual master list of holdings can be kept up to date rather simply by adding and replacing a few punched cards. Up-to-date listings can be prepared at any later date simply by running the revised deck of cards through the line printer.

Printed Book Catalogs and Catalog Cards. A complete list of library holdings is a natural outgrowth of the use of tab equipment to prepare lists of periodicals or other special collections. However, in this case, it is possible to use the manipulative properties of the cards to better advantage. Using one or more cards for each book in the library, it is possible to use the same deck of cards to prepare listings by author, title, subject category, call number, or other special category. Two examples of library catalogs prepared this way were given in Chapt. 2 (see Figs. 2-7 and 2-8). The book catalog, prepared by this, or other methods, has some distinct advantages over the standard card catalog; it is much more compact (due in large part to the fact that it contains less information), its contents can be searched and read in less time, and it is often easier to use. Its major advantage seems to be the fact that multiple copies can be prepared and distributed rather easily. The major disadvantage of a book catalog is that it is obsolete by the time it is printed, and gets progressively mpre out of date with time. The book-form catalog must thus be thought of as a throw-away item, to be replaced every month or so. The earliest application of tab equipment to the preparation of book catalogs was made in 1951.[16] The largest book catalog operation currently under way appears to be that of the Los Angeles County Library System, which publishes a book catalog for 117 branch libraries, with more than 540,000 entries in the subject catalog, and more than 150,000 entries in the author catalog.[17] Book catalogs of library holdings are currently used by several other organizations, several of which have prepared the catalogs with tab equipment.[18] Tab equipment has also been used to print multiple copies of conventional catalog cards (see Fig. 6-18), as well as weekly accessions bulletins.[19]

Machine Posting for Coordinate Indexing Systems. The task of posting new entries to a file of Uniterm

[14] McCann, A., "Applications of Machines to Library Techniques: Periodicals," ***American Documentation*****, Vol. 12, No. 4, pp. 260–265 (October 1961).**

Nicholson, N. N., and W. Thurston, "Serials and Journals in the M.I.T. Library," ***American Documentation*****, Vol. 9, No. 4, pp. 304–307 (October 1958).**

Anthony, L. J., and J. E. Hailstone, "Use of Punched Cards in Preparation of Lists of Periodicals," ***Association of Special Libraries and Information Bureaux Proceedings*** **(British), Vol. 12, No. 10, pp. 348-360 (October 1960).**

Logan, I. C., ***A Statistical Method for Evaluating Serial Holdings*****, an un-numbered report of the Utah State University, Logan, Utah (1961).**

[15] Berry, M. M., "Application of Punched Cards to Library Routines," Chapt. 13 in ***Punched Cards*****, Casey et al., editors, cited previously.**

[16] Alvord, D., "King County Public Library Does It with IBM," ***Pacific Northwest Libraries Association Quarterly*****, Vol. 16, No. 3, pp. 123–132 (April 1952).**

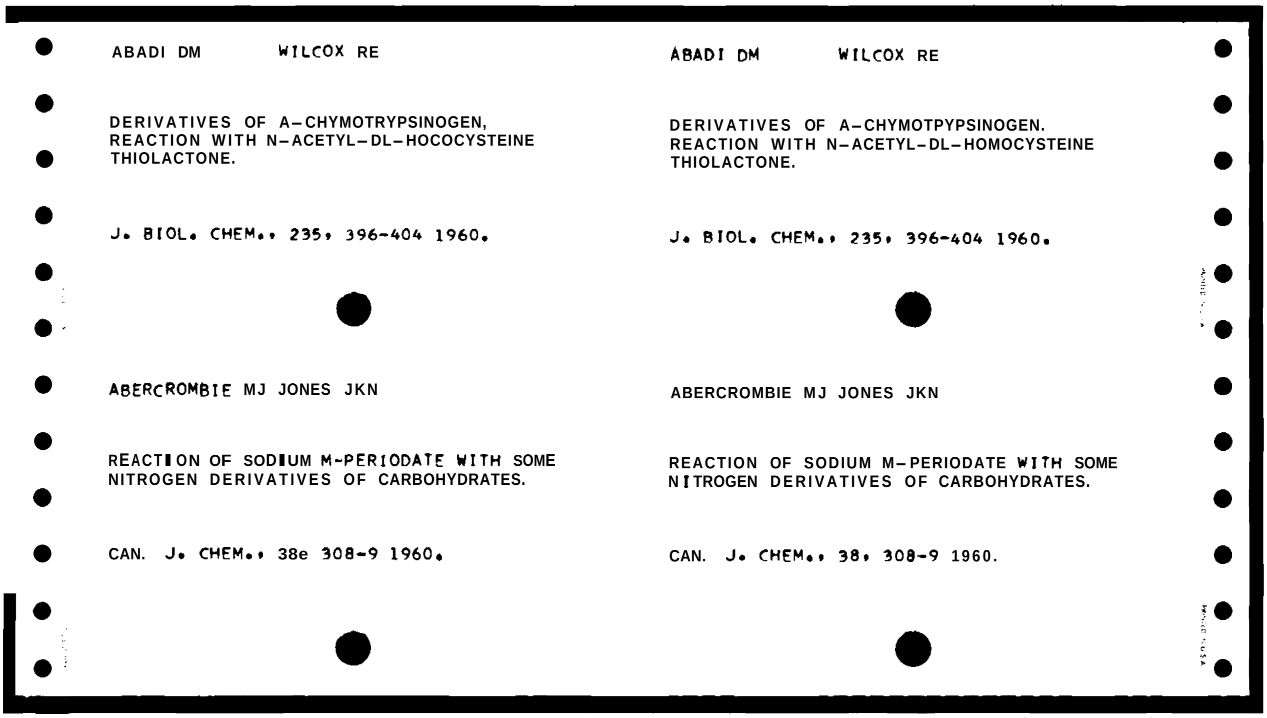
[17] MacQuarrie, C., "IBM Book Catalog," ***Library Journal*****, Vol. 82, No. 5, pp. 630–634 (March 1, 1957).**

[18] Vertanes, C. A., "Automation Raps at the Door of the Library Catalog," ***Special Libraries*****, Vol. 52, No. 5, pp. 237–242 (May-June 1961).**

Griffin, M., "Printed Book Catalogs," ***Revue de la Documentation*****, Vol. 28, No. 1, pp. 8–17 (February 1961).**

Dewey, H., "Punched Card Catalogs—Theory and Technique," ***American Documentation*****, Vol. 10, No. 1, pp. 36–50 (January 1959).**

[19] Durkin, R. E., and H. S. White, "Simultaneous Preparation of Library Catalogs for Manual and Machine Applications," ***Special Libraries*****, Vol. 52, No. 5, pp. 231–237 (May–June 1961).**


ABADI DM WILCOX RE

DERIVATIVES OF A-CHYMOTRYPSINOGEN,
REACTION WITH N-ACETYL-DL-HOCOCYSTEINE
THIOLACTONE.

J. BIOL. CHEM., 235, 396-404 1960.

ABERCROMBIE MJ JONES JKN

REACTION OF SODIUM M-PERIODATE WITH SOME
NITROGEN DERIVATIVES OF CARBOHYDRATES.

CAN. J. CHEM., 38e 308-9 1960.

ABADI DM WILCOX RE

DERIVATIVES OF A-CHYMOTPYPSINOGEN.
REACTION WITH N-ACETYL-DL-HOMOCYSTEINE
THIOLACTONE.

J. BIOL. CHEM., 235, 396-404 1960.

ABERCROMBIE MJ JONES JKN

REACTION OF SODIUM M-PERIODATE WITH SOME
NITROGEN DERIVATIVES OF CARBOHYDRATES.

CAN. J. CHEM., 38, 308-9 1960.

FIG. 6-18 Library catalog cards prepared by punched card equipment.

cards accounts for a significant fraction of the operating costs of a coordinate indexing system. Many posting schemes have been tried (e.g., typing, handwriting, numbering machines, and adhesive stickers) but the majority of installations continue to post with a typewriter.[20] One method for using tab equipment for Uniterm posting has been reported which claims a production cost of 60 to 75 dollars per 100 documents, which may be lower than the cost of doing it manually in many systems.[21] Their posting actually consists of the printing by the tab equipment of a new card for each term card that has been assigned additional accession numbers since the last machine run. At the end of a machine run, any card will show the cumulative list of accession numbers assigned to it up to that time. Machine posting has the additional advantage of easily providing a few additional copies of the Uniterm index. Machine posting appears especially attractive for organizations that must maintain duplicate indexes by periodic reprinting, by the use of copying equipment, or by the separate and parallel posting and updating of each index. The machine method does not make it economical to post the new entries as soon as they are received; entries must be accumulated or batched for posting, in order to make the best use of the machine time. The timeliness of the printed indexes will naturally depend upon how frequently the batches are run.

Preparation of Printed Indexes. Permutation or Key-Word-in-Context indexes, described in Chapt. 2, can be prepared by tab equipment, as demonstrated by Ohlman, Citron, and Hart, and by Veilleux.[22] In fact, preliminary studies seem to indicate that it may be much less expensive to use tab equip-

[20] Gull, C. D., "Posting for the Uniterm System of Coordinate Indexing," ***American Documentation,*** Vol. 7, No. 1, pp. 9–21 (January 1956).

[21] Weinstein, S. J., and R. J. Drozda, "Adaptation of Coordinate Indexing System to a General Literature and Patent File: Machine Posting," ***American Documentation,*** Vol. 10, No. 2, pp. 122–129 (April 1959).

[22] Citron, J., L. Hart, and H. Ohlman, ***A Permutation Zndez to the "Preprints of the International Conference on Scientific Information,"*** Report SP-44 of the System Development Corp., Santa Monica, California (November 1958).

Veilleux, M. P., "Permuted Title Word Indexing Procedures for a Man-Machine System," in ***Machine Indezing: Progress and Problems,*** pp. 77–111 (American University, Washington, D.C., 1961).

ment than computer equipment to prepare a KWIC index.[23] The index published by Ohlman et al. was prepared with the use of five different types of tab equipment: the key punch, verifier, sorter, reproducing punch, and printer.

In addition to the library book catalogs mentioned earlier, author and subject indexes to selected fields of technical literature have also been prepared with tab card equipment (see Fig. 2-9).[24] Some attention has also been given to the tab card production of citation indexes for scientific literature.[25]

Other Directories and Publications. In addition to the preparation of library catalogs and indexes, tab equipment has also been used in a few cases to prepare entire reports or documents. Usually these reports are in the form of tabulations such as parts lists for a complex assembly, or a tabulation of the technical characteristics of a collection of components.[26] However, under some circumstances, an English text report (e.g., a collection of abstracts) might even be prepared in this way.

There are also many applications of tab equipment to the maintenance and printing of lists of names for direct mailings. The cards may be coded with descriptive information about the individual addressee so that selected portions of a master file may be extracted for special mailings. Some of these tab card files for a company's own mailing list may hold as many as 600,000 names.[27] The firm of O. E. McIntyre, Inc., which makes a business of selling name lists, has tab card files for 36,000,000 names.[28]

Situations Where the Search of a Permanent File Is the Major Objective

Tab card equipment has been used by a great many people for file-searching applications. Most of these applications had permanent files of cards which were passed through sorting equipment when a search was to be made. In most cases, each file item was represented by a single card (i.e., a unit-record) which contained a coded representation of the subject analysis of that item. In applications such as these, the usual practice is to pass the entire file through the sorting equipment in order to select the items of interest.[29] This amounts to a serial scan of the entire file. In some instances it was possible to divide the file into sub-sections so that only a portion of the entire file would have to be examined. This would have the advantage of requiring less machine time.

A few systems use term cards instead of item cards. That is, separate tab cards, or decks of tab cards can be established for each descriptor or subject heading to record the serial numbers of the file items that are indexed by those descriptors. This is the tab card equivalent of the Uniterm terminal digit card, and one or more tab cards (term cards) are punched with the serial numbers of relevant file items just as one or more sheets of paper (term cards) may be typewritten with the serial numbers of relevant file items as in Fig. 2-1. In the tab sys-

[23] Ohlman, H., "Historical Development of Mechanical Indexing," unpublished paper presented at the Annual Convention of the American Documentation Institute, Berkeley, California, October 1960.

[24] Lipetz, B., "A Successful Application of Punched Cards in Subject Indexing," ***American Documentation,*** Vol. 11, No. 3, pp. 241–246 (July 1960).

Garfield, E., "The Preparation of Subject-Heading Lists by Automatic Punched Card Techniques," ***Journal of Documentation,*** Vol. 10, No. 1, pp. 1–10 (March 1954).

Garfield, E., "The Preparation of Printed Indexes by Automatic Punched Card Techniques," ***American Documentation,*** Vol. 6, No. 2, pp. 68–76 (April 1955).

[25] Lipetz, B., "Compilation of an Experimental Citation Index from Scientific Literature," ***American Documentation,*** Vol. 13, No. 3, pp. 251–266 (July 1962).

Garfield, E., "Citation Indexes for Science," ***Science,*** Vol. 122, No. 3159, pp. 108–111 (July 15, 1955).

[26] Santarelli, P. F., and R. F. Cunningham, "Data Processing Techniques for Parts Cataloging," paper presented at the 6th Meeting of the Technical Writers Institute, Rensselaer Polytechnic Institute, Troy, New York (June 11, 1958).

[27] Lewis, B., and D. Taylor, "Captive Plant with Competition," ***The Reporter of Direct Mail Advertising,*** September 1961.

[28] Anon., "Punched Cards Control the Huge 'McIntyre' Market," ***Management and Business Automation,*** Vol. 2, No. 5, pp. 16–19 (November 1959).

[29] Kirschner, S., "A Simple, Rapid System of Coding and Abstracting Chemical Literature Using Machine-Sorted Punched Cards," ***Journal of Chemical Education,*** Vol. 34, No. 8, pp. 403-405 (August 1957).

Gelberg, A., et al., "A Program Retrieval of Organic Structure Information Via Punched-Cards," ***Journal of Chemical Documentation,*** Vol. 2, No. 1, pp. 7–11 (January 1962).

Starker, L. N., and J. A. Cardero, "The Cyanamid Organic Structure Code and Search System," ***Journal of Chemical Documentation,*** Vol. 2, No. 1, pp. 12–18 (January 1962).

tem, as in the manual Uniterm system, pairs of term decks are examined jointly to determine which serial numbers exist on both decks. This joint examination for common numbers in two decks of tab cards can be done relatively simply by a standard collator (see Fig. 6-8). This term card approach has the advantage of generally requiring less machine time than serial scanning since the search operation now consists of the examination of a few relatively small decks of cards rather than the entire card file.[80]

For many types of file searching, especially those applications with questions that are structured in a relatively complex manner, the statistical sorter (see Fig. 6-11) can be used more effectively than a collator or single-column sorter.[31] The statistical sorter is also suited to the searching of tab cards which utilize superimposed coding. Unfortunately, this unit is generally more expensive than the single-column sorter or the collator, and less likely to be found in a general tab card installation or service bureau.

For file-searching purposes, a number of coding techniques may be used for putting the information on a tab card. Direct coding is common, with an item and its descriptors punched into the same card. As mentioned earlier, in Chapt. 4, it is possible to superimpose information into fixed fields of a tab card. However, this process usually generates some invalid punch combinations which cannot be interpreted meaningfully by most of the different models of tab equipment. Consequently, many types of tab equipment will not be of any use for the machine preparation or processing of tab cards that are punched with superimposed coding.

In most situations the files to be searched are relatively small, seldom more than 20,000 cards. As mentioned earlier, the equipment used for searching varies from file to file: the single-column sorters, the multi-column statistical sorter, the COMAC, the IBM Universal Card Scanner, and specially modified or developed sorters such as the Patent Office ILAS. The single-column sorter is the most commonly used machine, probably because of its relatively low cost and the possibility of using it part-time at an existing facility.

In most applications the selected cards are used to prepare (manually or mechanically) a very simple bibliography, possibly only a list of document numbers. In a few cases, where an abstract is typed or mimeographed onto the original cards,[32] a more comprehensive result is obtained. This is useful only if the cards can be copied conveniently and cheaply (by some photocopy process perhaps), or if the interested searcher is present when the cards are selected. In either case, return of the cards for further use is delayed by this extra handling. Most of the file-searching applications are characterized by low duty-cycles, so that the delay is not a severe limitation. Since very few installations conduct more than 5 or 10 searches per day, it is only seldom that anyone must wait to have a search made. Several installations have been mechanized to a further degree by using a pre-punched paper tape to generate abstracts of bibliographies after the document number is specified.

The existing installations have been developed for searching many different types of information, typically files of chemical structures and properties, technical reports, and patents.

[80] Whaley, F. R., "A Deep Index for Internal Technical Reports," Chapt. 21 in *Injormation Systems in Documentation,* J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

Whaley, F. R., "The Use of a Collator in an Inverted File Index," *Special Libraries,* Vol. 53, No. 2, pp. 65–73 (February 1962).

Whaley, F. R., "Operational Experience with Linde's Indexing and Retrieval System," in *General Injormation Manual: Information Retrieval Systems Conference, September 21–23, 1969, Poughkeepsie,* Brochure E20-8040 of the IBM Corp., Data Proceasing Div., White Plains, New York (1960).

[31] Southern, W. A., "Mechanized Processing and Retrieval of Bio-Medical Information," *Methods of Information in Medicine,* Vol. 1, No. 1, pp. 16–22 (January 1962).

Schulze, E. L., "An Application of Automation in the Library: Indexing Internal Reports," *Special Libraries,* Vol. 52, No. 2, pp. 63-67 (February 1961).

Maierson, A. T., and W. W. Howell, "Application of Standard Busineas Machine Punched-Card Equipment to Metallurgical Literature References," *American Documentation,* Vol. 4, No. l, pp. 3–13 (January 1953).

Garfield, E., "Preliminary Report on the Mechanical Analysis of Information by Use of the 101 Statistical Punched Card Machine," *American Documentation,* Vol. 5, No. 1, pp. 7–12 (January 1954).

Rockwell, H. E., R. L. Hayne, and E. Garfield, "A Unique System for Rapid Access to Large Volumes of Pharmacological Data; Application to Published Literature on Chlorpromazine," *Federation Proceedings,* Vol. 16, No. 3, pp. 726–731 (September 1957).

[32] Spiro, H., "IBM-Card Abstract-Index on Shaped Charge Research," pp. 397–403 in *Documentation in Action,* J. Shera et al., editors (Reinhold Publishing Corp., New York, 1956).

Application to Other Library Activities

A number of bookkeeping operations within a library can be handled with tab card equipment. Since 1936 many libraries have used tab equipment for book-charging systems, to keep track of the loaned items and to minimize the number of overdue items.[33] Tab equipment has also been used for magazine circulation systems to address and route the magazines to interested readers with a minimum of delay and clerical effort.[34] There is currently a considerable amount of interest in the application of data processing equipment to library activities.[35]

REPRESENTATIVE COSTS

Basic card stock is relatively inexpensive, costing 1 to 2 dollars per thousand cards, depending on the size, type, and color of the card stock; special features such as pre-scoring or special printing increase the cost. The pre-scored Port-a-Punch cards are slightly more expensive, running approximately 3 dollars per thousand. Card sets vary in price, depending upon the composition of the set, but they will generally be between 4 and 10 dollars per thousand.

The lease, purchase, and maintenance costs for tab card equipment from various manufacturers are described in Tables 6-1 and 6-2. Some equipment is listed in these tables that was not described in the text; however, their names are relatively descriptive, and for the most part they are modifications or extensions of the equipment described. The maintenance service is usually provided at no extra cost with the leased equipment, but is not provided with the purchased equipment. For purchased equipment, the customer usually has the option of either performing his own maintenance or contracting for it from the manufacturer at the rates quoted in the tables. Approximately 80 percent of all tab card equipment in use is leased equipment. For nearly all the purchased equipment, the maintenance is obtained by a contract arrangement from the manufacturer.

There may be many instances in which the lease or purchase of tab card equipment solely for their use as information systems cannot be economically justified. In such cases, the prospective user may be able, on a part-time basis, to use equipment which is already installed in the accounting or data processing departments of the user's organization. Under some circumstances it may be advantageous to use the equipment at a local service bureau or-

TABLE 6-1

Approximate Lease, Purchase, and Maintenance Costs for Some of the IBM Punched Card Equipment

Item	Monthly Rental	Purchase Price	Monthly Maintenance
010 Manual Card Punch	$ 10	$ 600	$ 1
024 Card Punch (numeric)	35	2,150	13
024 Card Punch (alphanumeric)	40	2,350	13
026 Printing Card Punch (numeric)	55	3,625	12
026 Printing Card Punch (alphanumeric)	60	3,825	13
046 Tape-to-Card Punch	140	5,700	30
047 Tapeto-Card Print Punch	160	6,500	31
056 Card Verifier	50	2,725	16
082 Sorter (450 cards/min.)	40	2,575	10
082 Smter (650 cards/min.)	55	2,600	16
083 Sorter (1000 cards/min.)	110	7,500	20
084 Sorter (2000 cards/min.)	250	10,000	33
087 Collator	215–270	11,400–14,000	23–40
101 Statistical Sorter (450 cards/min.) Mod. 1	500	24,000	63
101 Statistical Sorter (450 cards/min.) Mod. 2	245	14,000	43
108 Card Proving Machine	350	17,500	46
402.403 Accounting Machine	195–500	17,690–23,090	26–64
407 Accounting Machine	400–920	30,200–56,350	40–140
409 Accounting Machine	1,365–1,485	79,300–86,500	188–210
514 Reproducing Punch	70–140	5,510–6,380	28–43
519 Document Originating Machine	85–150	5,350–6,945	25–44
528 Accumulating Reproducer	210–440	15,000–17,800	32–87
557 Alphabetic Interpreter	130–305	8,200–15,340	32–67
9310 Universal Card Scanner	—	25,000–30,000	—
9900 Special Index Analyzer	—	27,000	—
— Port-A-Punch Kit	—	8–15	—

Notes:
1. The exact price will depend upon the number of accessories or optional features which are used.
2. The maintenance figures in the table are for equipment which is less than 3 years old.
3. The rental and maintenance figures are for single-shift operation. There is an additional charge for multiple-shift operation.

[33] Anon., "Punched Cards Solve Book Control Problems," *The Punched Card Annual,* 1955–56, The Punched Card Publishing Co., Detroit, *1955,* pp. *117–119.*

Birnbaum, H., *General Information Manual: IBM Circulation Control at Brooklyn College Library,* Brochure *E20-OW2* of the IBM Corp., Data Processing Div., White Plains, New York *(1960).*

Anon., *General Information Manual: IBM Tele-Processing in Circulation Control at Public Libraries,* Brochure *E20-0077* of the IBM Corp., Data Processing Div., White Plains, New York (1960).

[34] Richardson, W. H., "Circulation Control," *Special Libraries,* Vol. *51,* No. *9,* pp. 494–496 (November *1960).*

Booser, R. *J.,* "The Use of Data Processing Equipment for the Control and Circulation of Magazines," *Special Libraries,* Vol. *51,* No. 6, pp. 297–300 (July–August *1960).*

[35] Schultheias, *L.* A., et al., *Advanced Data Processing in the University Library* (Scarecrow Press, New York, *1962).*

TABLE 6-2

Lease, Purchase, and Maintenance Costs for Some of the Remington Rand (RR) and Other Punched Card Equipment

Item	Monthly	Purchase Price	Annual Service Charges
Alphabetic Punch (RR)	$ 40	$ 2,430	$ 98
Automatic Verifier (RR)	60	4,170	177
Interpreter (RR)	90	5,900	250
Reproducing Punch (RR)	125	7,640	324
Tag-to-Card Converter (RR)	185	12,600	535
Tape-&Card Converter (RR)	75	5,065	215
Electronic Sorter (RR)	85	5,600	280
Duplicate Card Detector (RR)	45	2,780	118
Collating Reproducer (RR)	170	12,155	516
High-speed Collator (RR)	125	10,000	500
Alphabetic Tabulator (RR)	285 to 495	20,160 to 29,200	860 to 1,240
VariTyper FOTOLIST Model 90	395	9,750	350
VariTyper FOTOLIST Model 270	700	17,850	500
Recordak LISTOMATIC Model 1-S	425	8,750	415
Recordak LISTOMATIC Model 1	750	14,750	415
Recordak LISTOMATIC Model 3	750	22,500	415
Friden COMPOS-O-LINE	625	16,375	480
Benson-Lehner COMAC Mark-2	N.A.	25,000 to 35,000	N.A.

Notes:

1. The exact price will depend on the number of accessories or optional features which are used.
2. The maintenance figures in the table are for equipment which is less than 3 years old.
3. The rental and maintenance figures are generally for single-shift operation. There may be an additional charge for multiple-shift operation.

ganization. The following rates are representative of service bureau charges for the use of tab equipment:

Equipment	Hourly Charge
IBM 026 Key Punch	$ 4.00
IBM 056 Key Verifier	4.00
IBM 083 Sorter	6.00
IBM 087 Alphabetic Collator	7.00
IBM 407 Accounting Machine	16.00
IBM 519 Reproducing Punch	6.50
IBM 557 Interpreter	5.50
IBM 101 Statistical Sorter	14.00

Additional References

Busa, R., "The Use of Punched Cards in Linguistic Analysis," Chapt. 16 in ***Punched Cads,*** 2nd ed., R. S. Casey et al., editors (Reinhold Publishing Corp., New York, 1958).

Dekker, J., "Interfixing of Descriptors for Row-by-Row Coded Punched Card Machine Literature Searching Systems," Chapt. 8 in ***Information Retrieval and Machine Translation,*** Part 1, A. Kent, editor (Interscience Publishers, New York, 1960).

Dietrich, E. V., "Machine Retrieval of Pharmacological Data," ***Science,*** Vol. 132, No. 3439, pp. 1556–1557 (November 25, 1960).

Edge, E. B., et al., "System for Indexing Research Reports Using a Punched Card Machine," ***American Documentation,*** Vol. 8, No. 4, pp. 275–283 (October 1957).

Faden, B. R., "Information Retrieval: Punched Card Equipment," ***Special Libraries,*** pp. 197–200 (May–June 1959).

Freed, R. N., "Machine Data Processing Systems for the Trial Lawyer," ***The Practical Lawyer,*** Vol. 6, No. 4, pp. 73–96 (April 1960).

Gage, R. P., "How Punch Cards Aid in Medical Research," in ***Progress Report in Chemical Literature Retrieval,*** G. L. Peakes et al., editors (Interscience Publishers, New York, 1957), pp. 165–172.

Graham, M. H., B. S. Rildenbrand, and B. H. Weil, "Indexing of Fuel and Lubricant Additives by Machine-Sorted Punched Cards," pp. 173–199 in ***Progress Report in Chemical Literature Retrieval,*** G. L. Peakes et al., editors (Interscience Publishers, New York, 1957).

Greme, M., "A Card Format for Reference Files in Information Processing," ***Communications of the Association for Computing Machinery,*** Vol. 4, No. 2, pp. 90–98 (February 1961).

IBM Corp., "Machine Functions," Form 22-8208-3 (1957).

Jones, W. S., and P. H. Butterfield, "A Technical Information Service Using Punched Cards for Indexing and Retrieval," pp. 57-68 in ***Progress Report in Chemical Literature Retrieval,*** G. L. Peakes et al., editors (Interscience Publishers, New York, 1957).

Kuentzel, L., "New Codes for Hollerith-Type Punched Cards to Sort Infrared Absorption and Chemical Structure Data," ***Analytical Chemistry,*** Vol. 23, No. 10, pp. 1413–1418 (October 1951).

Livingston, G. A., and I. D. Welt, "Chemical Structures and Responses of Organisms to Applied Chemicals," Chapt. 16 in ***Information Systems in Documentation,*** J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

Luhn, H. P., ***General Rules for Creating Machinable Records for Libraries and Special Reference Files,*** Report L-419 of the IBM Corp., Advanced Systems Development Div., Yorktown Heights, N.Y. (September 30, 1959).

MacKinnon, F. B., J. C. Leary, and D. Levinson, "An Analysis of the Problem of an Experimental Adaptation of Punched Cards and Mechanical Retrieval to Legal Research and Indexing of State Statutes, Codes, and Session Laws," Chapt. 18 in ***Information Systems in Documentation,*** J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

Maloney, C. J., "A Remington-Rand Punched Card Data Retrieval System," ***American Documentation,*** Vol. 9, No. 1, pp. 1–12 (January 1958).

Patterson, H. A., et al., "A System for Context Storage and Retrieval of Information from the Published Literature, Applicable to Both the IBM 101 and Electronic Computers," in ***General Information Manual: Information Retrieval Systems Conference, September 21–93, 1969, Poughkeepsie,*** Brochure E20-8040 of the IBM Corp., Data Processing Div., White Plains, N.Y. (1960).

Peakes, G. L., "Experience with the Unit Card System for Report Indexing," Chapt. 19 in ***Information Systems in Documentation,*** J. H. Shera et al., editora (Interscience Publishers, New York, 1957).

Remington Rand, ***Functions of Remington Rand Univac Data Processing Systems,*** Brochure U1363, rev. 2 (1959).

Scheele, M., ***Punch-Card Methods in Research and Documentation with Special Reference to Biology*** (Interscience Publishers, New York, 1961).

Stiles, H. E., "Identification of the Conditions for Valid Application of Machines to Bibliographic Control," ***American Documentation,*** Vol. 9, No. 1, p. 42 (January 1958).

Weil, B. H., and E. A. Clapp, "The Adaptation of the ASM-SLA Metallurgical Literature Classification for Use in a Machine-Sorted Punched Card System," Chapt. 20 in ***Information Systems in Documentation,*** J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

Weil, B. H., and B. Hildenbrand, "Information on Fuel and Lubricant Additives from Machine-Sorted Punched Cards," Chapt. 17 in ***Information Systems in Documentation, J. H.*** Shera et al., editora (Interscience Publishers, New York, 1957).

- The Equipment
- The Programing
- Application to Information Processing and Retrieval
- Representative Costs

7

Computer systems

THE EQUIPMENT

Computer equipment that used electronic rather than mechanical assemblies was first developed and put into practical use in the early 1950's. After a few years of successful application to scientific and engineering calculations, computers began to be applied to problems of a commercial or business nature. Even though the equipment was initially designed for complicated arithmetic tasks rather than the handling of large volumes of alphanumeric data it did perform well enough to indicate that it would be feasible to use computers for tasks other than scientific calculations. In the past few years, the computer equipment field has become very competitive, and the computers have continuously shown great improvements in speed, flexibility, and suitsbility for a number of application areas.[1] Over 250 different models of computers have been built to date,[2] and over 13,000 computers were in use in 1962 in the United States?[3] There is every indication that equipment performance will continue to improve, and that new applications will become practical, such as information retrieval and text processing.

These computers are very complex pieces of electronic equipment. Fortunately, it is not necessary to know how to build one in order to be able to use one—just as it is not necessary to know how to build an automobile in order to be able to drive one. However, a rudimentary understanding of the basic workings of the computer can provide a better appreciation of it, and some insight into the types of problems that might be appropriate for computer application.

In many ways, the computer can be considered to be an extension of punched card equipment. It can do all the things that tab equipment can do, such as sorting, selecting, merging, and performing arithmetic operations. However, it can generally do it at much greater speeds, and with greater volumes of data. In brief, computer equipment might be thought of as grown-up tab equipment. The capabilities and potential that were first displayed by tab equipment many years ago are now seen in a more fully developed form in computer equipment. For example, the computer can do the same simple series of additions and multiplications that were first done on the tab equipment, but it can do them 1000 to 10,000 times faster. The computer also has a much larger memory or information storage capacity than tab equipment. Similarly, the computer is able to take in information and write out information much faster. The sequence of machine operations for each step of the job is still estab-

[1] Bourne, C. P., and D. F. Ford, "The Historical Development and Predicted State of the Art of the General Purpose Digital Computer," *Proceedings of the 1960 Western Joint Computer Conference* (Institute of Radio Engineers, New York).

[2] *Data Processing Equipment Encyclopedia: Electronic Devices* (Gille Associates, Detroit, Michigan, 1981).

Weik, M. H., *A Third Survey of Domestic Electronic Digital Computing Systems,* Report 1115 of the Ballistic Research Laboratories at Aberdeen Proving Ground, Maryland (March 1961), PB-171 265, AD-263 212.

[3] Anon., "JD and A Computer Census Results," *Automatic Data Processing Service Newsletter,* Vol. 7, No. 17 (January 21, 1963).

Anon., "Computer Census as of January 1963," *Business Automation,* January 1963, pp. 42–43.

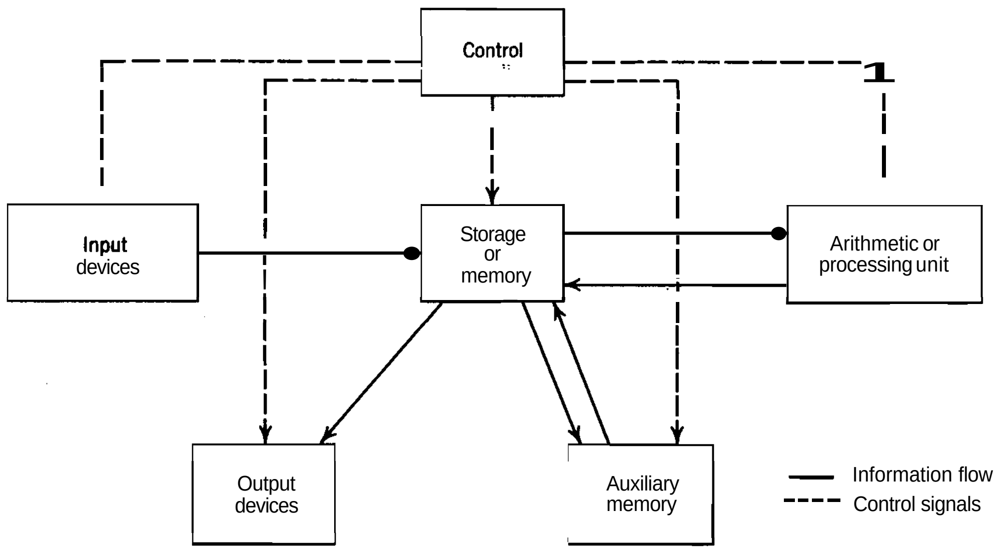


FIG. 7-1 Computer block diagram.

lished initially by a human, although the computer does have the capability to change its given list of instructions in a specified manner during the execution of the program. The reader of this book who is generally unfamiliar with computers may find it useful to think of them as high-speed tab card systems which are controlled by written lists of sequential machine steps ("programs") rather than by plugboards, and which have all the features of regular tab equipment, but in an expanded and more powerful form. For a more complete explanation of some of the terms used in this chapter, the reader is referred to the American Institute of Physics glossary.[4] In a strict pedantic sense, a computer could be defined as a device for manipulating numbers very rapidly; or a device which can accept information, process it, and supply the results of the processing. The processing generally involves performing arithmetical and logical operations on the information. Actually, the computer field has grown so fast, and in so many directions, that it has been difficult to get agreement on a definition of a computer. When the word "computer" is used in this book it will always refer to digital computers rather than analog computers; and more specifically, it will refer only to general-purpose computers rather than special-purpose computers which have been developed primarily for application to a single and very specific problem. As mentioned earlier, the term "digital" means that all of the operations are carried out by using discrete numbers, and all of the data are stored as discrete numbers.

The exact size, shape, and characteristics of general-purpose digital computers vary widely, however most of the computers built to date have a number of common features. Most computers have the same general structure—an arithmetic or processing unit, a control unit, a storage or memory unit, and input-output equipment for man-machine communication. Figure 7-1 shows the functional relationship of these units in block diagram form. Each of these blocks is described briefly in the following paragraphs.

The computer control unit interprets the instructions given to it by the operator, and then schedules and directs the activities of all the other electrical circuits and units of equipment in order to get the job done. It acts as the foreman.

The arithmetic unit is the mill or processor which, under the leadership of the control unit, operates on the information given to it and performs all of the logical testing and processing required. It can perform arithmetic operations such as addition and multiplication, as well as many logical operations such as comparisons, and the extractions of selected portions of a string of numbers or characters.

[4] Fox, P., *Glossary of Terms Frequently Used in Physics and Computers* (American Institute of Physics, New York, 1962, 22 pp.).

The memory or storage units retain the necessary information (initial data, intermediate results, final results, and the program) in a form that can be interpreted by the control unit. The kind of storage device most often used for the main computer memory is the magnetic core since they can be operated at much higher speeds than many other types of memory devices. A sample of an array of such cores is shown in Fig. 7-2. It is these core planes that store the information inside the computers, and there may be several hundred such planes in a single large computer. However, the fast core memories are relatively expensive, and a compromise between cost and speed is often obtained by augmenting the main core memory with a slower but less expensive auxiliary memory.

Most auxiliary memory devices such as magnetic tape require a serial or sequential access to the recorded information rather than a random access. Reading information from a magnetic tape, for example, requires that some tape be passed over before the desired information can be located. Consequently, the time required to locate a particular item depends upon where it is located on the tape, and how much of the file material must be passed over in order to reach it. For random interrogations of the file, such serial access memory is relatively inefficient. Any of several types of random access storage devices may be more efficient than serial access devices for random interrogations. These random access devices, such as the main core memories and magnetic disc files, permit any item in the memory to be read or written in about the same amount of time, which is usually much shorter than the average time required with a serial access device such as magnetic tape. Examples of magnetic drum and disc files are shown in Figs. 7-3 and 7-4, respectively. A close-up of the discs and the selection mechanism is shown in Fig. 7-5. The term "random" does not refer to the way in which the information is organized in the storage, nor does it imply immediate access. The term means that any memory location, chosen randomly, requires about the same amount of access time as any other memory location. Many existing systems such as the RAMAC disc, the tape bins, and some drum systems, are compromises between serial and immediate access memories. For permanent storage of large amounts of information (e.g., personnel and inventory records, documents lists) magnetic tape is a convenient medium. The tapes containing the information can be removed and stored for later processing, since the recorded information can be played back into the computer. The main advantages of magnetic tape are its very high storage capacity (as much as 50 million digits per reel for some computers) and its very high reading or writing rates (more than 100,000 digits per second for some computers). An example of a representative

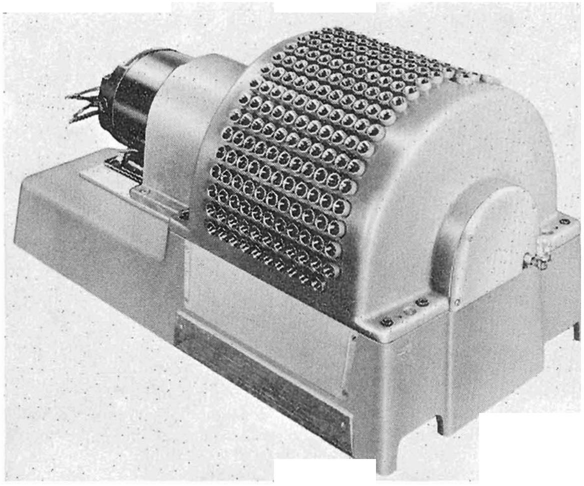


FIG. 7-3 Sperry Rand magnetic drum memory unit.

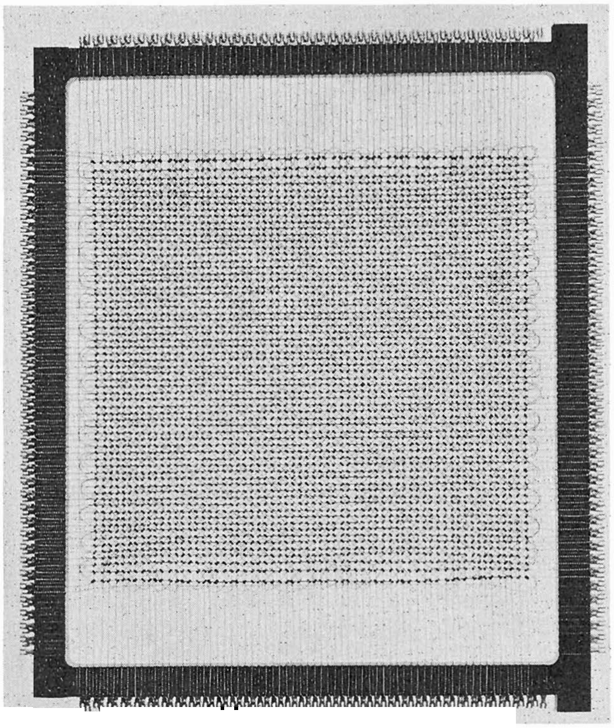


FIG. 7-2 Portion of a magnetic core memory unit.

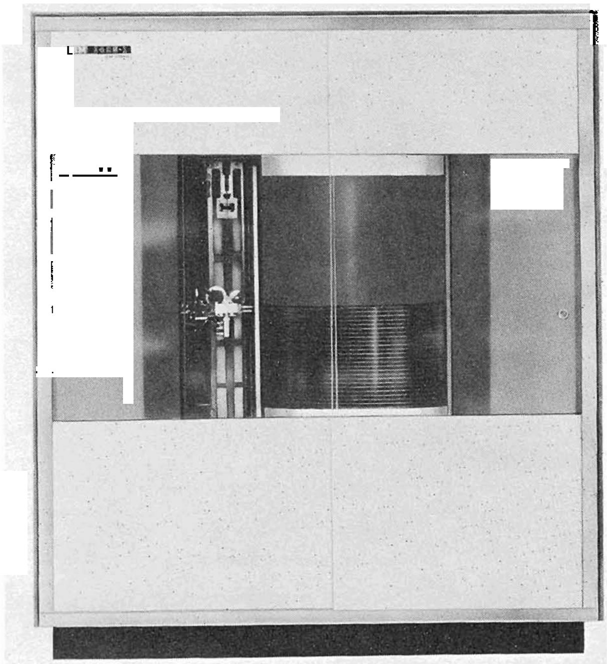


FIG. 7-4 IBM 1405 magnetic disc storage unit for the IBM 1401 data processing system.

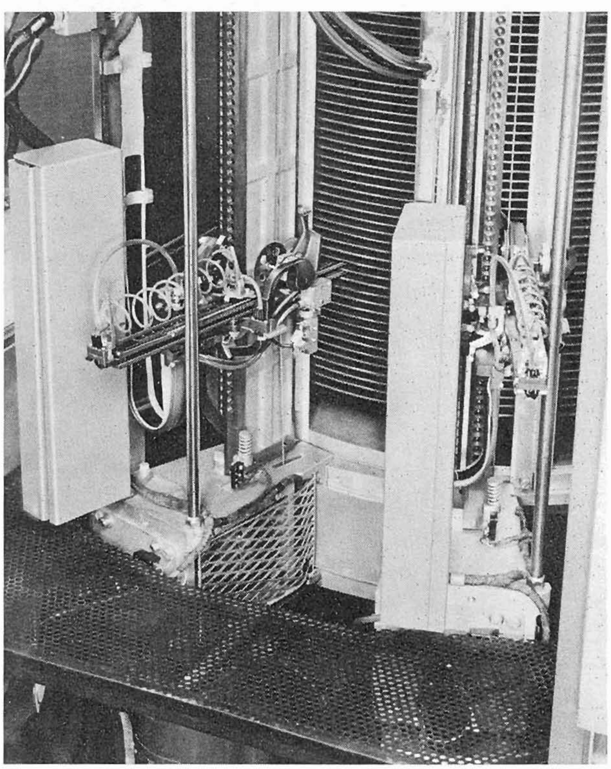


FIG. 7-5 Close-up of dual access arms on the disc memory unit for the IBM RAMAC 305.

magnetic tape unit for a computer is shown in Fig. 7-6.

Means must be provided for the machine and the user to communicate with each other. In most cases this communication should be rapid enough to ensure that the computer does not have to wait for data or instructions to be transferred in or out of the machine. Information can be fed into the computers at a relatively high rate of speed by any of several devices, such as punched card readers, magnetic tape transports, paper tape readers, and character recognition equipment. Most of the input equipment is so designed that information recorded

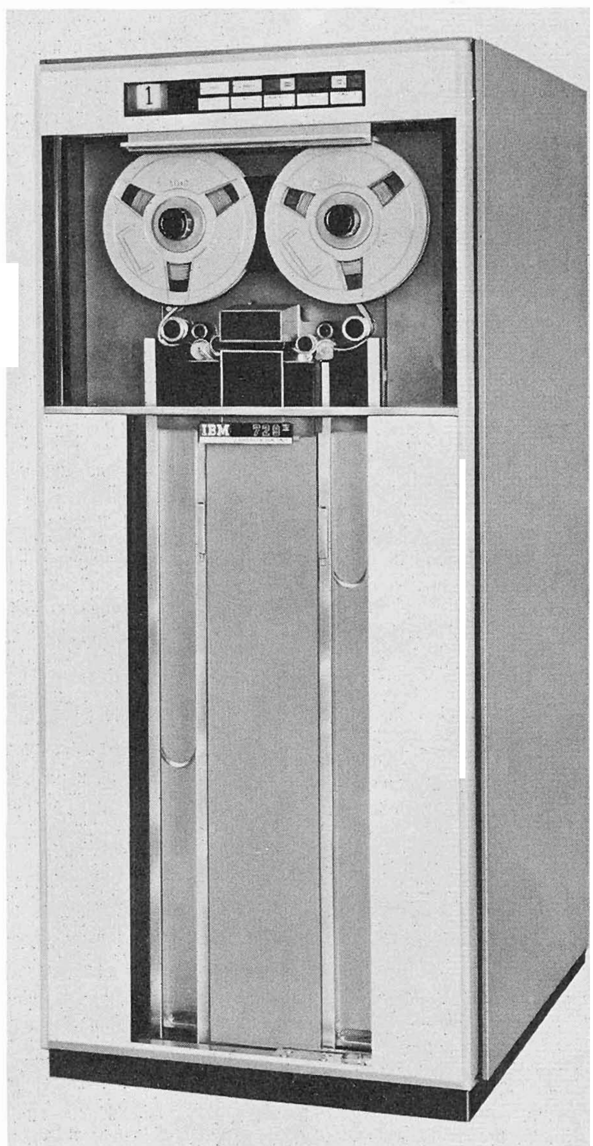


FIG. 7-6 IBM 729-4 magnetic tape transport.

at action points (cash registers, typing stations, etc.) can later be read directly into the computer without an additional transcription process.

Several types of output device are used with computers: line printers (150 lines per minute is common, but speeds go as high as 5000 lines per minute), card punches, and paper tape punches are most common.

Illustrations of some representative computer systems are given in Figs. 7-7 through 7-9. Some of the units such as the magnetic tape units, line printers, punched card readers, and operator consoles are easily recognized in these illustrations. The large enclosed cabinets contain most of the electronic circuitry of the computer system, and may house the control, arithmetic, and some memory units, most of which have no recognizable physical form as such. Operator and maintenance consoles or displays are used on every computer to indicate the status of various critical parts of the system, and to permit some degree of manual control and intervention as required.

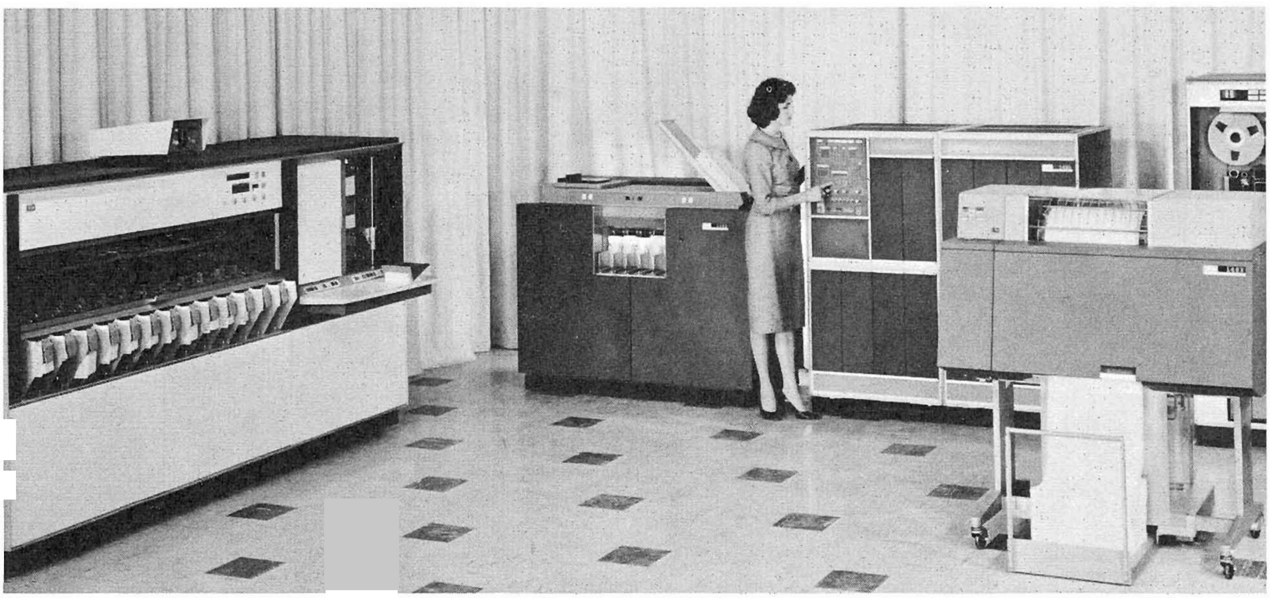


FIG. 7-7 IBM 1401 data processing system with Model 1418 optical character reader.

THE PROGRAMING

A computer must be given detailed instructions for every task that it is to perform—just as a complex piece of tab card equipment must be given a specified sequence of machine steps by the plugboard wiring. The computer is a very fast mechanical moron that can perform only a limited number of minor operations. Any credit for cleverness, ingenuity, or sophistication in the computer's performance belongs to the man who prepared the list of instructions (sequence of operations) for the computer. The writing of these detailed lists of instructions is called *programing*, and a computer's ability to play chess or to search for documents is no better than the programer's ability to write specific instructions for these operations from a fixed repertoire of machine operations and to foresee all situations that the computer might have to consider. The actual list of steps or machine operations is called a program. This repertoire of standard instructions is different for every model of computer, and the programs or results obtained by a programer depend to a large measure on how well he can describe the particular job and formulate the machine's operations in terms of the programing language.

For most computers, a program is a list of separate instructions, each of which directs the machine to perform some elemental task. The instruction is usually a defined set of numbers or charac-

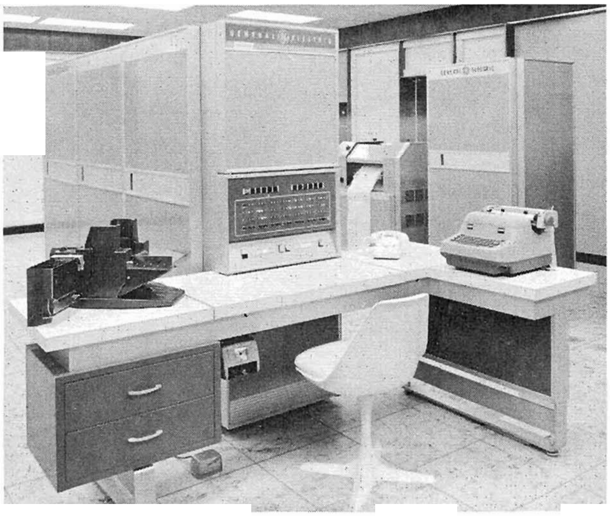


FIG. *7-8* General Electric Model 225 computer.

ters arranged in a certain manner, with each number having a predetermined meaning that has been built into the control circuits of the computer. For example, .for the IBM 7070, the instruction +5100219999 is interpreted by the computer as meaning "Test tape/disc storage channel No. 2 to see if it is busy; if it is busy, get the next instruction from memory location 9999; if it is not busy, execute the next instruction." For the same computer the instruction +8100551200 is interpreted as meaning "Write, with zero suppression, on magnetic tape transport No. 5 on channel No. 1, the records whose locations are defined by the contents of memory location 1200." Programs written in this very basic notation (machine language) are easy for the computer to interpret, but are difficult for the programer to read and write. Special programs (automatic programing aids) are available for most computers that will act as mechanical translation programs to translate simpler human statements into more detailed machine language. These automatic programing aids ("assemblers," "compilers," "interpreters") relieve the programer, but often require additional computer time to generate or run the program. The FORTRAN, ALGOL, FACT, and COBOL programing systems are representative of automatic programing techniques for general applications.[5]

[5] Acronyms for FORmula TRANslator (FORTRAN), ALGOrithmic Language (ALGOL), Fully Automatic Compiling Technique (FACT), and Common Business Oriented Language (COBOL).

These programing systems allow the programer to write his programs in a much more natural language, such as the following examples of a portion of a program whose statements were written in FACT, an automatic compiling technique for the Honeywell H-800 computer:

REPLACE NAME, STREET, CITY, STATE, TELNO, AND RATE OF MASTER BY NAME, STREET, CITY, STATE, TELNO, AND RATE OF DETAIL.

SORT E ON G AND F WITHIN C AND H ON D AND I WITHIN A.

These statements, written by the programer, are translated by the computer program into the long sequence of detailed machine instructions required to do these tasks. Although most of the automatic programing techniques were designed to be used with a particular computer, much work has been done recently to develop problem-oriented languages that can be used by almost any computer. Programs written in ALGOL, for example, can be treated almost as universal programs in the sense that they can be run on almost any large computer, using special translating programs provided by each manufacturer for his particular machine.

The writing of an extensive program is done in about the same way that an extensive report is written; a crude and then perhaps a more detailed outline is made to show what is to be done, and then the detailed writing is done to fit into the structure of the outline. In a manner similar to inserting

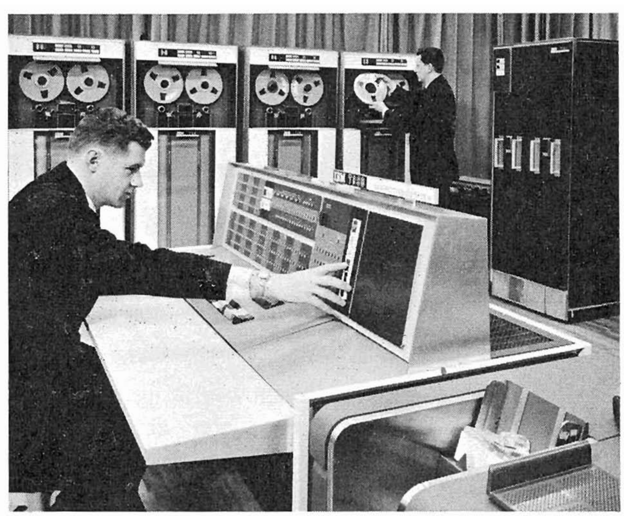


FIG. *7-9* *IBM 7090* data processing system.

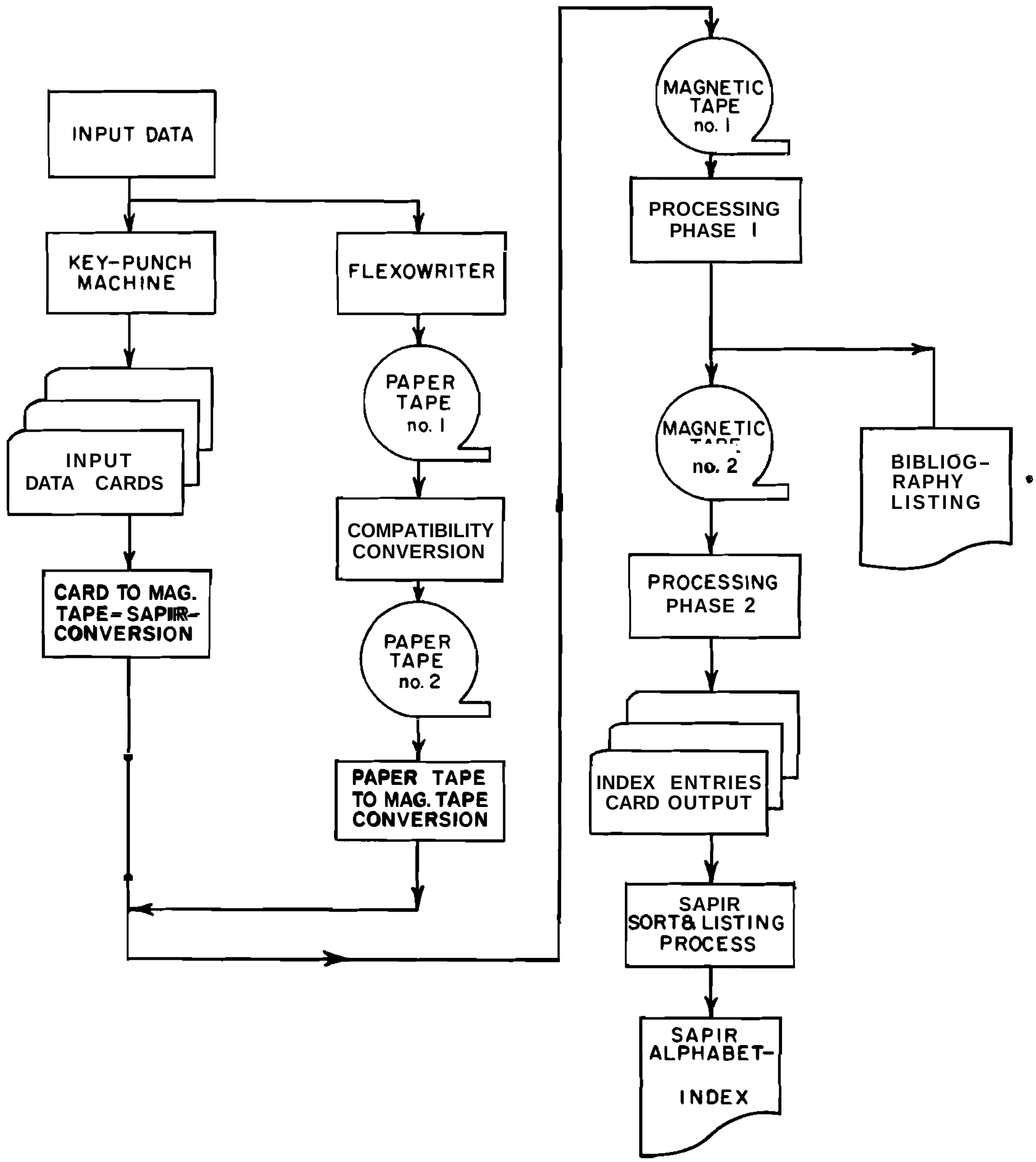


FIG. 7-10 Basic flow chart for a Key-Word-in-Context program.

standard paragraphs in a new report, some standard programs (often called "sub-routines") may be inserted into the skeleton or framework of the new program. This approach is well illustrated by Figs. 7-10 through 7-12, which show the general flow charts (the coarse outline), the detailed flow charts (the detailed outline), and the instruction list (the text) for a program written to perform KWIC indexing with an IBM 1401 computer.[6] The flow charts provide a two-dimensional array or display, to better illustrate the operations and interrelationships of the different parts of the program

[6] **Turner, L. D., and J. H. Kennedy,** System of Automatic Processing and *Indexing* of Reports, **Report UCRL-6510, University of California, Lawrence Radiation Laboratory, Livermore, California (July 1961).** **Giallanza, F. V., and J. H. Kennedy,** Key-Word-In-Title (KWIT) *Index* for Reports, **Report UCRL-6782 of the University of California, Lawrence Radiation Laboratory, Livermore, California (May 14, 1962).**

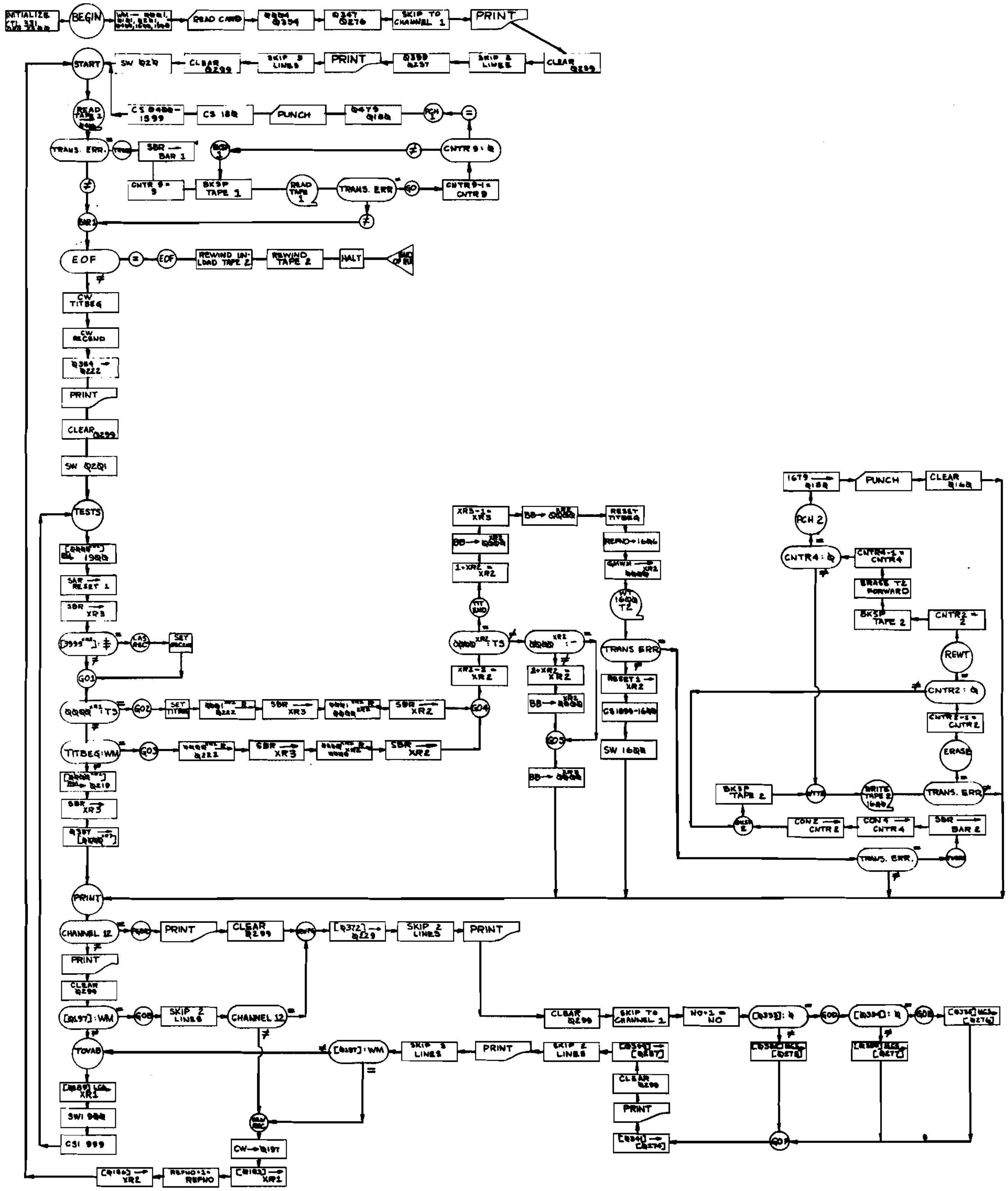


FIG. 7-11 Detailed flow chart for a portion of a Key-Word-in-Context program.

```
CLEAR STORAGE 1    ,008015,022026,030034,041,045,053,0570731026
CLEAR STORAGE 2    L072116,110106,1051178101/I99,027A075029)027B0010270B026/0991,001/001117I00
BOOTSTRAP CARD     ~008015~022029~0560631056029                                 ,0240671056

PG LIN  CT  LABEL   OP    A OPERAND          8 OPERAND     0      LOC    INSTRUCTION COMMENTS   KWIC1

1 010               CTL   331
1 020               ORG   2500
1 030                     • KEY-WORD-IN-CONTEXT INDEX IBM 1401 DATA
1 040                       PROCESSING PROGRAM BY LESTER O. TURNER, 041061
1 050   7   BEGIN   SW    0001               0101                 2500   , 001 101
1 060   7           SW    0201               0400                 2507   , 201 400
1 070   7           SW    1600               1900                 2514   , W00 Z00
1 080   1           R                                             2521   1
1 090   7           MCW   0004               0354                 2522   M 004 354
1 100   7           MCW   0347               0276                 2529   M 347 276
1 110   2           CC                                     1      2536   F 1
1 120   1           W                                             2538   2
1 130   4           CS    0299                                    2539   / 299
1 140   2           CC                                     K      2543   F K
1 150   7           MCW   0399               0257                 2545   M 399 257
1 160   1           W                                             2552   2
1 165   4           CS    0299                                    2553   1 299
1 166   4           SW    0201                                    2557   , 201
1 170   2           CC                                     L      2561   F L
1 180   8   START   MCW   (UI                0400          R      2563   M (U) 400 R
1 181   5           8     TRERR                            L      257'I  B P10 L
1 182   4           CW    TITBEG                                  2576   ) 198
1 184   4           CW    RECENO                                  2580   I 1 9 7
1 200   5           B     EOF                              K      2584   8 P78 K
2 010   7           MCW   0354               0222                 2589   M 354 222
2 020   1           W                                             2596   2
2 030   4           CS    0299                                    2597   1 299
2 040   7   TESTS   MCM   0000        1      1900                 2601   P 0+0 ZOO
2 050   4           SAR   RESET1                                  2608   Q 189
2 060   4           SBR   XR3                                     2612   H 099
2 070   7           A     CON10M             0100                 2616   A M95 100
2 075   8           B     LASREC             0000       3         2623   8 Q17 0+0
2 080   8   GO1     B     GO2                0000       1         2631   8 025 0+0
2 090   8           BWZ   GO3                TITBEG            I  2639   V Q99 198 1
2 100   7           MCM   0000        1      0219                 2647   P 020 219
2 110   4           SBR   XR3                                     2654   H 099
2 120   7           MCW   0337               0000       3         2658   M 337 0+0
2 130   5   PRINT   ▮     PGENO                            -      2665   8 A20
2 140   1           W                                             2670   2
2 150   4           CS    0299                                    2671   1 299
2 160   8           BWZ   GCB                RECENO            1  2675   V B63 197 1
2 170   7   TOVAB   MCW   0189               XR1                  2683   M I89 089
3 020   4           CS    1999                                    2690   / Z99
3 030   4           SW    1900                                    2694   , ZOO
3 032   4           CS    1999                                    2698   / Z99
3 034   4           SW    1900                                    2702   , ZOO
3 040   4           ▮     TESTS                                   2706   8 001
3 050   4   TRERR   SBR   BAR1   +003                             2710   H P42
3 060   7           MCW   CON9               CNTR9                2714   M 190 358
```

FIG. 7-12 Portion of a listing of the **instructions** of a **Key-Word-in-Context** program.

than would be obtained by the program listing, which is a linear array. When documented to this degree of detail, this program could be utilized directly by another organization with the same type of computer or programing language.

As mentioned earlier, some automatic programing techniques are available to simplify the programing effort. However, many information retrieval or information processing tasks cannot conveniently use these techniques in their present form, and many workers are developing automatic programing techniques which are more suitable to these problems.[7]

A considerable amount of intellectual effort is

[7] Banes, A. V., H. L. Engel, and D. R. Swanson, "An Instruction Code for Language Processing," Western Electronic News, November 1959, pp. 44-48.

Gelernter, H., et al., "A Fortran-Compiled List-Processing Language," Journal of the Association for *Computing* Machinery (*ACM*), Vol. 7, No. 2, pp. 87–101 (April 1960).

Kehl, W. B., et al., "An Information Retrieval Language for Legal Studies," Communications of the ACM, Vol. 4, No. 9, pp. 380–389 (September 1961).

Hoffman, J., and A. Opler, "Use of MOBL in Preparing Retrieval Programs," Communications of the ACM, Vol. 4, No. 9, pp. 389-391 (September 1961).

required to write and check the programs for large data processing systems. Many commercial data processing programs have required dozens of man-years of programing effort for completion. Military data processing installations such as the air defense systems have taken as many as 100 man-years of programing effort to develop certain operational capabilities. In addition to this initial programing effort, it is not uncommon for large computer facilities to maintain staffs of 10 to 50 people to make new programs and modify existing ones. A considerable amount of effort is also required to do the preliminary study and determination of data processing requirements. A recent book describes how such a study was done for the University of Illinois library.[8]

It takes time to produce a program that operates correctly. Very rarely does a program work correctly the first time it is given to a computer. A series of test runs on the computer is usually necessary to locate the programing errors and ensure that the program is operating satisfactorily. Once the program is working correctly, it may be used again at any time in the future, with assurance that it will run without trouble. In theory, once the initial programing investment has been made, the program can be run again at any time in the future, with no additional cost beyond the machine operating cost. However, this is not always the case, since many organizations are continually forced to modify finished programs because of such factors as changes in procedures, policies, and data formats. Once a program is running satisfactorily, it should be documented and distributed so that other users can benefit from the work. Such cooperation (akin to the cooperation that has long existed in regard to other technical literature) will enable computer users to build up libraries of programs or sub-routines that can be used at any time in the future for a computer of the same model and equipment complement. The programing language, internal machine language, and repertoire of instructions are different for most models of computer, so that programs written for one computer generally cannot be used on a different model or a computer from a different manufacturer. Table 7-1 lists some computer programs for information retrieval that are available from some of the equipment manufacturers and user groups.

APPLICATION TO INFORMATION PROCESSING AND RETRIEVAL

A given problem can generally be handled by any of several different models of computers, and need not be given to a specific computer type. A literature-searching system or KWIC index system, for example, can be implemented with almost any model of computer. The fact that most of the past and current literature on these two applications describes an implementation in terms of one or two of the present models of a single prominent manufacturer should not obscure the fact that almost any other type of computer could also have been used. As a matter of fact, literature searching systems have been implemented with at least a dozen different types of computers. The following sections discuss several basic types of information handling and retrieval problems as implemented with several different types of computers.

The reader should be reminded again that the use of a computer with an information retrieval system will not necessarily enhance or improve the efficiency of retrieval — that is something that will depend primarily upon the indexing and classification techniques that are used. The presence of computer equipment per se cannot provide better retrieval or guarantee good performance although it might provide the answers in a shorter period of time, or at

Saurmet, J. E., H. Ohlman, and H. G. Bohnert, "Discussion—The Pros and Cons of a Special IR Language," *Communications of the ACM,* Vol. 5, No. 1, pp. 8–10 (January 1962).

Colilla, R. A., and B. H. Sams, "Information Structures for Processing and Retrieving," *Communications of the ACM,* Vol. 5, No. 1, pp. 11–16 (January 1962).

Yngve, V. H., "COMIT as an IR Language," *Communications of the ACM,* Vol. 5, No. 1, pp. 19–28 (January 1962).

Barnes, R. F., "Language Problems Posed by Heavily Structured Data," *Communicatwns of the ACM,* Vol. 5, No. 1, pp. 28-34 (January 1962).

Cheatham, T. E., Jr., and S. Warshall, "Translation of Retrieval Requests Couched in a 'Semiformal' English-Like Language," *Communications of the ACM,* Vol. 5, No. 1, pp. 34-39 (January 1962).

Sable, J. D., "Use of Semantic Structure in Information Systems," *Communicatwns of the ACM,* Vol. 5, No. 1, pp. 40–43 (January 1962).

Grems, M., "A Survey of Languages and Systems for Information Retrieval," *Communications of the ACM,* Vol. 5, No. 1, pp. 43-46 (January 1962).

Kellog, C., "The Fact Compiler: A System for the Extraction, Storage and Retrieval of Information," *Proceedings of the 1960 Western Joint Computer Conference,* pp. 73–82, Institute of Radio Engineers, New York.

[8] Schultheiss, L. A., et al., *Advanced Data Processing in the University Library* (Scarecrow Press, New York, 1962).

TABLE 7-1

Available Computer Programs for Information Retrieval and Related Activities

1. "IBM 1401–KWIC (Key Word in Context)." Card version available from IBM, Los Angeles, Education Center. Tape version under development for submission to 1401 Library.
2. "IBM 1401–SDI-3 (Selective Dissemination of Information)." Developed by Advanced Systems Development Division (ASDD) and submitted to 1401 Library. Identification (ID) number 10.3.004.
3. "IBM 1401–INFORM." A complete information retrieval system developed jointly by IBM and North American Aviation and submitted to the 1401 Library. ID number not known yet.
4. "IBM 1401–SNOOP." File search using superimposed coding submitted to 1401 Library. ID number not known yet.
5. "IBM 1401–Inverted File Program." File look-up with a card system, submitted to the 1401 Library. ID number not known yet.
6. "IBM 7090–Permuted Title Index," also known as BE-PIP. Developed by Bell Telephone Laboratories and available from SHARE Library as SDA No. 1239.*
7. "IBM 7090–KWIC." Developed by IBM, Owego, and submitted to the SHARE Library. ID number not known yet.
8. "IBM 7090–SDI-2." Developed by IBM–ASDD and submitted to the SHARE Library. ID number not yet known.
9. "IBM 7090–Literature Retrieval." Under development by IBM, San Jose Research Laboratories.
10. "GE-225 Electronic Search Program." Available from General Electric Co.
11. "IBM 650–Information Retrieval System," J. T. Ahlin, in General *Information Manual: Information Retrieval* Systems Conferences, September 21–23, 1969, *Poughkeepsie,* Brochure E20-8040 of the IBM Corp., Data Processing Div., White Plains, New York.
12. "Automatic Information Retrieval Program for the IBM-650," No. 12.0.007 of the Program Information Dept., IBM Corp., White Plains, New York. This is a search of a coordinate index on punched card discs, with 14 document numbers per card.
13. L. D. Turner and J. H. Kennedy, "System of Automatic Processing and Indexing of Reports," Report UCRL-6510, University of California, Lawrence Radiation Laboratory, Livermore, California (July 1961).

*SHARE Distribution Agency, c/o IBM Corp., 590 Madison Avenue, New York 22, New York.

less cost. As mentioned earlier, the retrieval performance (i.e., relevancy and recall) will depend primarily upon the intellectual organization of the index, and not upon the method of implementation.

Analysis of Text Structure

Concordances. A considerable amount of effort has been expended on the mechanization of the production or assembly of information. The semi-automatic preparation of indexes, abstracts, and concordances is an outgrowth of this effort. The preparation of concordances is one writing task that has almost gone over completely to preparation by computers.

Concordances are printed alphabetical lists that indicate the places in the text of a body of work where particular words were used. They are very useful in studies of language, vocabulary, and the history of ideas in literature and philosophy, and have been used primarily for that purpose (i.e., textual criticism). In addition to showing the location where the word was used, the concordance may also show a small amount of the text material that precedes and follows the particular words at each location. If the adjacent text material is not shown, then the concordance simply amounts to a detailed index to all of the words used in the text. If the location of the word is not shown, then the concordance simply amounts to a word list (i.e., a complete list of all the different words used in the text). Such a concordance might be constructed to serve as the first draft of a detailed index to a corpus of text material. It might also be used to provide a list of terms for the subsequent manual preparation of a thesaurus or dictionary.

The first concordance to the Bible was published by John Marbeck in 1550, and no one knows how many years it took to compile. An exhaustive concordance, published by James Strong in 1894, required 30 years of effort. In 1955, a 2000-page concordance of the Bible produced with a Univac computer required approximately 150 hours of computer time after several months of planning and preparation.[9] The entire Revised Standard Version of the Bible, more than 800,000 words, was put on 4 reels of magnetic tape. With the exception of 132 com-

[9] McCulley, W. R., "Univac Compiles a Complete Bible Concordance," Systems, Vol. 20, No. 2, pp. 22–23 (March–April 1956).

mon words, the stored words were arranged alphabetically along with their context and location in the Bible—book, chapter, and verse. The printed concordance contained approximately 350,000 contexts on about 2000 pages.

Some other early work in mechanized concordance preparation, initially with tabulating equipment but later with computers, was done by a group of Italian scholars and theologians who prepared a complete concordance (1.6 million punched cards) to the Summa Theologica by St. Thomas Aquinas.[10]

In 1959, a 965-page concordance to the poems of Matthew Arnold, prepared by data processing equipment, required approximately 40 hours of IBM 704 time.[11] The total body of Arnold's published poetry consists of about 17,000 lines. The complete body of the concordance was formated and printed by the computer, and the pages were then photographed and a book published by the photo-offset process. No typesetting was required. The study group that prepared the Matthew Arnold concordance states that this is the first of a series of concordances to be prepared with data processing equipment for publication by Cornell University Press. An example from one of their concordances in preparation, *A* Concordance to the Poem of W. *B.* Yeats, is shown in Fig. 7-13. Another concordance being prepared at Cornell is a concordance to The Anglo-Saxon Poetic Records, a series published in 6 volumes (about 168,000 words of text). This work is also being done on IBM 704 equipment.[12] Such a concordance is often useful for textual critics who look for small differences between subsequent printings or different versions of the same work, and also look to see how the author might repeat himself in different works.

Other concordance applications include a study of the Dead Sea Scrolls,[13] a study of the language used in legal statutes,[14] and studies of the derivation and use of words in folk songs.[15] The use of computing equipment for concordance construction [16] has shortened the elapsed time required for such a task, and has permitted scholars to devote more of their attention to the analysis and interpretation functions, and less to clerical chores. In that regard, it has been a useful tool for linguistic and textual research. The main problem, so far, is the practical one of getting the text material into machine-readable form. This problem is complex—but by no means insurmountable.

Dictionaries and Thesauri. Any computer program that does some processing of the file information (e.g., file searching or computer preparation and printing of KWIC indexes, accession lists, or catalog cards) is in a position to also provide some useful by-product information such as a list of words used in the file and their frequency of use or association with other words in the file. One of the products of the computer run for each issue of the KWIC index to Biological Abstracts, for example, is an alphabetical list of the "junk" words that the program has been instructed to ignore, as well as a count of the number of times that each of these terms and the key words chosen were used in that issue. Two particular by-products that might be prepared with the assistance of computer programs are a dictionary and a thesaurus.

The dictionary, as we normally consider it, is a list of words and their associated word roots, as well as a definition of that word group. In machine translation work, the dictionary is usually constructed as a table to show corresponding words or

[10] Tasman, P., *"Literary Data Processing," IBM Journal, Vol.* 1, *No.* 3, *pp.* 249–256 *(July* 1957).

[11] *Painter, J. A., "Computer Preparation of a Poetry Concordance," Communications of the Association for Computing Machinery, Vol.* 3, *No.* 2, *pp.* 91–95 *(February* 1960).

Parris, *S. M., editor, A Concordance to the Poems of Matthew Arnold* (Cornell *University Press, Ithaca, N.Y.,* 1959).

[12] Bessinger, *J. B., "Computer Techniques for an Old English Concordance," American Documentation, Vol.* 12, *No.* 3, *pp.* 227–229 *(July* 1961).

[13] Tasman, *P., Indexing the Dead Sea Scrolls by Electronic Literary Data Processing Methods, an* un-numbered *report of the* IBM *World Trade* Corp. *(November* 1958).

[14] Kehl, *W. B., et al., "An Information Retrieval Language for Legal Studies," Communications of the Association for Computing Machinery,* Vol. 4, *No.* 9, *pp.* 380–389 *(September* 1961).

[15] Sebeok, *T. A., and V. J. Zeps, "Computer Research in* Psycholinguistics: *Toward an Analysis of Poetic Language," Behavioral Science,* Vol. 6, *No. 4, pp.* 365–369 *(October* 1961).

[16] *Current Research and Development in Scientific Documentation, No.* 9, *Report* NSF-61-76 *of the National Science Foundation, Washington, D.C. (November* 1961).

Wisbey, R., "Concordance Making by Electronic Computer: Some Experiences with the 'Wiener Genesis,'" Modern Langauge Review, Vol. 57, *No.* 2, *pp.* 161–172 *(April* 1962).

Lamb, S. M., "The *Digital Computer as an Aid in Linguistics," Language, Vol.* 37, *No. 3, pp.* 382–412 (1961).

```
                                                                PAGE     TITLE          LINE
MAGI   (CONTINUED)
    OF THE CROWNED MAGI: AND THE HOUND OF CU  . . . . . . .    169 SECRET ROSE     V    9
    OF THE CROWNED MAGI: AND THE KINGS WHOSE EYES . . . . .    169 SECRET ROSE     V    9
    THE MAGI  . . . . . . . . . . . . . . . . . . . . . . .    318 MAGI                 T
MAGIC
    WERE'T NOT THAT THERE IS MAGIC IN HIS HARP,  . . . . . .   222 SHADOW WATER B      16
    THAT THERE IS MAGIC IN THAT HARP OF HIS,  . . . . . . .    222 SHADOW WATER B V    16
    A MAGIC THAT CAN CALL A DEMON UP, . . . . . . . . . . .    238 SHADOW WATER B     344
    I WILL END ALL YOUR MAGIC ON THE INSTANT. . . . . . . .    242 SHADOW WATER B     411
    BY MAGIC STRINGS, I'LL MAKE THIS ANSWER TO IT:  . . . .    243 SHADOW WATER B     440
    MY MAGIC STRINGS, I'LL MAKE THIS ANSWER TO IT:  . . . .    243 SHADOW WATER B V   440
    IF YOU HAD TAKEN ME BY MAGIC SPELLS, . . . . . . . . . .   245 SHADOW WATER B     490
    WHAT MAGIC DRUM?  . . . . . . . . . . . . . . . . . . .    559 MAGIC DRUM           T
    THROUGH LIGHT-OBLITERATING GARDEN FOLIAGE WHAT MAGIC DRUM?  560 MAGIC DRUM           4
    TAKEN IN MAGIC, IN THE CHURCH'S NAME . . . . . . . . . .   694 MOSADA 1           104
    WITH PITILESS MAGIC HAS BOUND YOU--  . . . . . . . . . .   T08 FAIRY PEDANT        31
    SO MEN IN LIFE AND I IN MAGIC PLAY:  . . . . . . . . . .   T22 WITCH VIVIEN        55
MAGICAL
    CHAUNT IN HIS EAR DELUSIONS MAGICAL, . . . . . . . . . .   111 CUCHULAIN SEA       80
    GO, CAST ON HIM DELUSIONS MAGICAL,  . . . . . . . . . .    111 CUCHULAIN SEA   V   80
    THE MAGICAL POWERS TO AND FRO. . . . . . . . . . . . . .   138 TO IRELAND      V   24
    MAGICAL SHAPES.  BY THE HELP OF AN IMAGE  . . . . . . .    367 EGO DOMINUS          7
    OBEDIENT TO SOME HIDDEN MAGICAL BREATH. . . . . . . . .    382 VISION ROBART       13
    MAGICAL UNICORNS BEAR LADIES ON THEIR BACKS. . . . . . .   426 PHANTOM HATRED      18
MAGICIAN
    OBEYS.  SOME FIERCE MAGICIAN FLIES OR WALKS  . . . . . .   72D WITCH VIVIEN         9
MAGNANIMITIES
    THEIR MAGNANIMITIES OF SOUND, . . . . . . . . . . . . .    404 FOR DAUGHTER        44
MAGNANIMITY
    THE STRENGTH THAT GIVES OUR BLOOD AND STATE MAGNANIMITY     481 BLOOD AND MOON      29
       . . .
    AS THOUGH WITH MAGNANIMITY OF LIGHT, . . . . . . . . . .   618 BRONZE HEAD          9
MAGNIFICENCE
    MONUMENTS OF ITS OWN MAGNIFICENCE;  . . . . . . . . . .    40T SAIL BYZANTIUM      14
    LOWLY THE HEAD OF HIS MAGNIFICENCE  . . . . . . . . . .    689 NETTLESHIP           6
    DOWN LOW THE HEAD OF HIS MAGNIFICENCE  . . . . . . . . .   689 NETTLESHIP      V    6
MAGNIFIED
    MY MEMORIES HAD MAGNIFIED  . . . . . . . . . . . . . . .   399 BREAK OF DAY        10
MAGNIFY
    CONSIDER MOST TO MAGNIFY, OR TO BLESS, . . . . . . . . .   418 ANCEST HOUSES       39
MAGNITUDE
    THAT MADE THE MAGNITUDE AND GLORY THERE . . . . . . . .    483 VERONIC NAPKIN       5
MAGNUS
    OF MAGNUS ANNUS AT THE SPRING, . . . . . . . . . . . . .   43T SONGS PLAY 1         T
MAHOGANY
    LIT UP THE DARK MAHOGANY AND THE WINE, . . . . . . . . .   410 TOWER               27
MAHRAJAS
    OF RAJAS AND MAHRAJAS BEYOND NUMBER? . . . . . . . . . .   T24 KANVA ON SELF        6
MAID      SEE WAITING-MAID SERVING-MAID
    AND NOW A MAID, ON A SWIFT BROWN STEED  . . . . . . . .     12 OISIN 1         V  142
    MAID NIAM FROM A LITTLE TRUMP  . . . . . . . . . . . . .    14 OISIN 1         V  194
    AND KNELT THEM, EVERY MAID AND MAN. . . . . . . . . . .     15 OISIN 1         V  214
    THEY HUSHED THEM, EVERY MAN AND MAID.  . . . . . . . . .    1T OISIN 1         V  235
    WITH ONE LONG GLANCE ON MAID AND BOY . . . . . . . . . .    19 OISIN 1         V  28T
    AROSE FROM EVERY MAID AND BOY, . . . . . . . . . . . . .    20 OISIN 1         V  289
    HE SPAKE TO THE YOUNG MAN, "IS THERE NO MAID  . . . . .    109 CUCHULAIN SEA       58
    THEN CONCHUBAR SENT THAT SWEET-THROATED MAID, . . . . .    110 CUCHULAIN SEA       72
    THEN CONCOBAR SENT THAT SWEET THROATED MAID  . . . . . .   110 CUCHULAIN SEA   V   72
    AND LIFT YOUR TENDER EYELIDS, MAID, . . . . . . . . . .    125 GOES FERGUS          5
    MAID QUIET  . . . . . . . . . . . . . . . . . . . . . .    171 MAID QUIET           T
    WHERE HAS MAID QUIET GONE TO, . . . . . . . . . . . . .    171 MAID QUIET           1
    I NEVER HAVE SEEN MAID QUIET  . . . . . . . . . . . . .    171 MAID QUIET      V    1
    NO WAITING MAID SHOULD EVER SPREAD  . . . . . . . . . .    194 BAILE AILLINN   V  120
    A SERVING MAID AND A SERVING MAN, . . . . . . . . . . .    301 RUN PARADISE    V   13
    AND EVERY MAN AND MAID AND BOY . . . . . . . . . . . . .   337 TOM OROUGHLEY        2
    WHETHER TO MAID OR HAG:  . . . . . . . . . . . . . . . .   525 DANCING DAYS        22
    "I AM SORRY," SAID THE MAID,  . . . . . . . . . . . . .    595 COLONEL MARTIN      29
    TWO MEN WHO LOVE ONE MAID HAVE AMPLE CAUSE . . . . . . .   659 ISLE STAT II 1      30
    FOR THIS THING FIGHT--WE LOVE ONE MAID?  HER NAME? . . .   661 ISLE STAT II 1      66
    THE PLAIN, A LITTLE FAIR-HAIRED MAID I LOVE. . . . . . .   679 ISLE STAT II 3     312
    THERE DWELLS A LITTLE FAIR-HAIRED MAID I LOVE.  . . . .    679 ISLE STAT II 3 V   312
    MY LORD, YOU CALLED?  NOT I.  THIS MAID IS DEAD. . . . .   704 MOSADA 3           125
    THEN EACH WIFE AND MAID AND MOTHER CALLS BY NAME           719 PHANTOM SHIP        24
       SOME DROWNED ONE.
    DO NOT FEAR US, EARTHLY MAID!  . . . . . . . . . . . . .   726 LOVERS QUARREL       1
    TEASING EVERY WILFUL MAID. . . . . . . . . . . . . . . .   727 LOVERS QUARREL      28
MAIDEN
    A MAIDEN, ON A SLENDER STEED,  . . . . . . . . . . . . .     2 OISIN 1         V   7V
    "D PLEASANT MAIDEN," ANSWERED FINN, . . . . . . . . . .      5 OISIN 1         V   40
    NOW, PLEASANT MAIDEN, TELL TO ME  . . . . . . . . . . .      5 OISIN 1         V   43
    OF THE YOUNG.  "YOUNG MAIDEN, WHAT MAY BRING . . . . . .     5 OISIN 1         V   50
    AND NOW A MAIDEN RODE LIKE THE WIND  . . . . . . . . . .    12 OISIN 1         V  142
    AND ONCE A MAIDEN BY MY SIDE  . . . . . . . . . . . . .     16 OISIN 1         V  231
    FULL MANY MORE, WHEN PEERED A MAIDEN SWEET . . . . . . .    33 OISIN 2         V  68F
    A MAIDEN WITH SOFT EYES LIKE FUNERAL TAPERS, . . . . . .    33 OISIN 2         V   69
    OH MAIDEN."  "ARE YE SPIRITS OF THE SEA,  . . . . . . .     34 OISIN 2         V  T8B
    THAT MAIDEN FOUND A RING HUNG ON THE WALL. . . . . . . .    37 OISIN 2         V  120
    AND AH, YOU PROUD MAIDEN, YOU ARE NOT SO FAIR WHEN HIS OAR   91 OLD FISHERMAN .      9
    PROUD MAIDEN, YE ARE NOT SO FAIR, WHEN HIS OAR  . . . .     91 OLD FISHERMAN   V    9
    NO MAIDEN LOVES ME, NO MAN SEEKS MY HELP, . . . . . . .    103 FERGUS DRUID    V   26
    MAIDEN, COME FORTH: THE WOODS KEEP WATCH FOR THEE: . . .   645 ISLE STAT I 1        1
```

FIG. 7-13 Portion of the *Concordance of the Poems of W. B. Yeats* prepared by computer equipment.

word roots in some second language for every word or word root used in the primary language. The computer preparation of a dictionary is an operation similar to concordance construction. However, for dictionary work, there is less emphasis on where the words were used, and more emphasis on what words and word roots were used. The computer programs cannot do the entire task of preparing the dictionaries, but they can compile an alphabetical list of all the words encountered to date, or all the words that have not yet been entered into the dictionary. This at least removes much of the burden of clerical work in the preparation of the dictionary. However, a considerable amount of human effort is still necessary to provide the definitions or secondary language equivalent.

Most of the dictionary compilation work with computers has been done with the research programs in machine translation.[17] In fact, every machine translation programing system incorporates some specific programs to generate and maintain a dictionary, and to point out new words as they are encountered for the first time. Such programs may be helpful for the development of reference retrieval tools.[18] Several other studies have suggested the use of computer techniques for the construction of glossaries and similar works.[19]

The main use of thesauri has already been discussed in Chapt. 2. However, it will be useful at this point to describe again, what is meant by the word "thesaurus." A thesaurus is a collection of words (e.g., synonyms) classified by concept or broad clusters of meaning, rather than by an alphabetic or other characteristic. In documentation work, it is sometimes used to designate a schedule of authorized subject headings, arranged so that like words are grouped together and appropriate "see" and "see also" references are specified (see Fig. 2-5). A considerable amount of effort is being expended by such organizations as ASTIA, the American Institute of Chemical Engineers, and the Engineers Joint Council, to develop and maintain thesauri for their subject fields of interest.

The thesaurus may be used as an authority list for indexing or classification work in order to maintain a degree of control over the choice and use of indexing terms.[20] It may also be used to help suggest terms to be used for search requests. Swanson, for example, has compiled and used a thesaurus as a tool in some studies of computer indexing and searching of natural text material.[21] In these studies, the thesaurus was compiled to perform the machine indexing and to assist in the machine searching experiments. Each entry in the thesaurus was weighted to reflect the importance of that word for retrieval purposes. The thesaurus was used to automatically formulate a series of search instructions for a natural language question and to compute the relevance scores for documents retrieved.

As with the dictionary compilation, the thesaurus cannot be compiled entirely by the computer. However, the computer can be used to perform much of the clerical work such as the initial sorting and listing of all words that make reference to every other word. The programs could also locate all the dead end references—that is, a reference that is made *to* a particular word, but not made back *from*

[17] The status of the machine translation work in mid-1960 is well described in the report, "Research on Machine Translation," Hearings before the Special Investigating Subcommittee, Committee on Science and Astronautics, U.S. House of Representatives, 86th Congress, 2nd Session (U.S. Govt. Printing Office, May 1960).

Oettinger, A. G., et al., "Linguistic and Machine Methods for Compiling and Updating the Harvard Automatic Dictionary," *Proceedings of the International Conference on Scientific Information,* Vol. 2, pp. 951–973 (National Academy of Sciences, Washington, D.C., 1959).

[18] Howerton, P. W., "The Application of Modern Lexicographic Techniques to Machine Indexing," pp. 326–330 in *Machine Indexing: Progress and Problems* (American University, Washington, D.C., 1961).

[19] Booth, A. D., and A. J. T. Colin, "On the Efficiency of a New Method of Dictionary Construction," *Information and Control,* Vol. 3, No. 4, pp. 327–334 (December 1960).

Colin, A. J. T., "The Automatic Construction of a Glossary," *Information and Control,* Vol. 3, No. 3, pp. 211–230 (September 1960).

Levison, M., "The Application of the Ferranti-Mercury Computer to Linguistic Problems," *Information and Control,* Vol. 3, No. 3, pp. 231–247 (September 1960).

[20] ASTIA, *Automation of ASTIA—1960,* pp. 4–9, AD-247 000, Armed Services Technical Information Agency, Arlington, Virginia (December 1960).

ASTIA, *Automation of ASTIA—1959. A Preliminary Report,* pp. 43–50, AD-227 000, Armed Services Technical Information Agency, Arlington, Virginia (December 1959).

See also fn. 22 in Chapt. 2.

[21] Swanson, D. R., "Interrogating a Computer in Natural Language," *Information Processing 1962,* pp. 288–293, C. M. Popplewell, editor (North Holland Publishing Co., Amsterdam, 1963).

Swanson, D. R., "Research Procedures for Automatic Indexing," pp. 281–304 in *Machine Indexing: Progress and Problems* (American University, Washington, D.C., 1961).

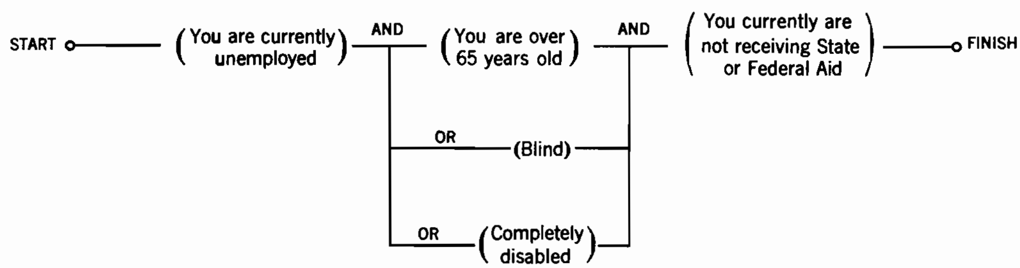
Swanson, D. R., "Searching Natural Language Text by Computer," *Science,* Vol. 132, No. 3434, pp. 1099–1104 (October 21, 1960).

Swanson, D. R., see previous reference in Chapt. 2, fn. 24.

that word. This first draft assembly and checking of terms by the computer can be a very helpful start for the human effort which must follow to arbitrate which terms should be placed where, and what generic and coordinate relationship they should have to the other terms in the thesaurus.

Text Schematics. Some interesting techniques have been developed for analyzing legal statutes, business contracts, and other complex text material by breaking them down into schematic logic diagrams to show their basic meaning.[22] A very simple example of this conceptual simplification is given by the following fictitious statute:

You are entitled to County unemployment benefits if: [23]



It would seem unnecessary to employ machine techniques to analyze a statute as simple as this example. However, as everyone knows, there are many statutes that are virtually incomprehensible, often with single sentences that run for a page or more. Kehl reported one instance of a Pennsylvania statute with a sentence that was 692 words long.[24] Information processing techniques are currently available that might be able to make sense of these incomprehensible statutes, without the help of a lawyer. Consider a semi-automatic editing system in which legal statutes drafted for legislative bodies are first submitted to a machine editor to generate logic diagrams of the form illustrated below. The logic diagram could then serve as a "check print" for the person who drafted the legislation, to make sure that the text actually says what the draftsman really intended to say (or, if the legislator wasn't quite sure what he had in mind in the first place, the logic diagram might help him pull his thoughts together). Once the thought and the formulator of the thought were in agreement, the diagram could accompany the text and be published with the text so that anyone who had occasion to read the statute would have it presented in a very concise and unambiguous manner.

It may well be that commercial publishers will soon find it profitable to publish manuals of diagrams such as these to accompany the publication of laws or discourses on active fields of legislation—particularly tax legislation and the rulings of the Federal regulatory commissions. Anyone who has ever tried to penetrate the farther reaches of these legal jungles ought to be happy to pay a fair amount of money for a moderately good map.

It is unreasonable to expect complete coverage of all of our statutes in the near future; the magnitude of the task is staggering. It is estimated that, in 1960 alone, approximately 14,000,000 words of Federal statutes were published. However, machine analysis of legal statutes seems to hold promise for the future. And if machine analysis can simplify ambiguous or needlessly complex legal documents, it could also help with other types of documents.

The same techniques, for example, may also be applied to point out the ambiguities or flaws that may exist in the business documents and messages of any large business or government organizations.[25]

[22] Allen, L. E., "Toward a Procedure for Detecting and Controlling Syntactic Ambiguity in Legal Discourse," Chapt. 42 in *Information Retrieval and Machine Translation,* Part 2 (Interscience Publishers, New York, 1961).

[23] If, that is, you can successfully pass through the network. The statute in text form might read, "You are entitled to County unemployment benefits if you are currently unemployed and if, in addition, you are over 65 years old, or blind, or completely disabled, and if, in addition, you are not currently receiving State or Federal Aid."

[24] Kehl, article cited earlier in this chapter, fn. 7.

[25] Allen, L. E., "Toward More Clarity in Business Communications by Modern Logical Methods," *Management Science,* Vol. 5, No. 1, pp. 121–135 (October 1958).

Although these particular techniques have not been tested to date with any computer programs, it seems quite feasible to do so. The approach is similar to the computer work that is currently being done on syntax studies of English text material.

Preparation of Retrieval Tools

Indexing, Cataloging, and Classification. Chapter 2 has already introduced the subject-indexing problem and mentioned some of the progress made in a number of areas of computer indexing. So far, the major practical application of computers to indexing has been along the lines of the KWIC index preparation, with KWIC publications such as conference proceedings, special collections of literature, and lists of their library's serial or periodical holdings prepared by at least 30 organizations. Several technical journals are also using machine techniques to semi-automatically assemble and compose title bulletins for publication.[26] Various figures have been quoted for the cost of preparing a KWIC index with computer equipment; ranging from 25 cents to 1 dollar per bibliographic item to prepare the master copy of the KWIC index. Examples of representative KWIC indexes have been illustrated in Figs. 2-2 and 2-3. Equipment parts catalogs have also been prepared with computer equipment.[27]

One recent example of the use of a computer to prepare a subject index is the recently initiated publication, *Chemical-Biological Activities,* a computer-prepared index to the current literature on the biological activity of organic compounds.[28] Relevant information is extracted from a number of primary journals and prepared for input to an IBM 1401 computer. The computer processes the information to prepare six different indexes to that particular collection of information. A good illustration of the types of indexes that are prepared is shown in Fig. 7-14. For this particular printout, the IBM 1403 printer has been modified to provide both uppercase and lowercase letters, as well as some special symbols.

Another good example of computer assistance in the production of an index is the work currently being done by the Institute for Scientific Information to prepare the *Index Chemicus,* a periodical listing of new chemical compounds along with the pertinent references to show where they were cited in the current literature. The current *Index Chemicus* and its cumulative index (*Encyclopaedia Chimica Internationalis*) are prepared with IBM 7090 and 1401 computers and consist of 4 special indexes: a molecular formula index, a permuted formula (RotaForm) index, an author index, and a journal index. The first 6 volumes and 4500 pages of this encyclopaedia covered over 17,000 articles and approximately 180,000 chemical compounds. The rotated formula index of this encyclopaedia is a permuted index which forms all possible written sequences of elements for a given compound, and lists each compound under each of the several possible formulas that might be used to describe it.

Figure 7-15 shows a portion of the computer printout for a citation index of the type currently being developed by the Institute for Scientific Information. It is expected that the data base for this experimental index will eventually include some 3,000,000 citations from about 500 scientific journals, primarily for the field of genetics, and will show the citation linkages of these articles. In the particular example shown, references are listed in alphabetic order by author. Each of these source citations is followed by the citation for the paper that did the citing. This index then points to continuations of, or use made of, the original work in other studies. The first draft of this index consisted of approximately 8000 pages of computer printout in a form similar to the sample page shown.[29]

[26] Markus, J., "State of the Art of Published Indexes," *American Documentation,* Vol. 13, No. 1, pp. 15–30 (January 1962).

Kennedy, R. A., "Library Applications of Permutation Indexing," *Journal of Chemical Documentation,* Vol. 2, No. 3, pp. 181–185 (July 1962).

General Information Manual: Key-Word-in-Context (KWIC) *Indexing,* Brochure E20-2091 of IBM Corp., Data Processing Div., White Plains, New York (1962).

Giallanza, F. V., and J. H. Kennedy, *Key-Word-in-Title* (KWIT) *Index for Reports,* Report UCRL-6782 of the University of California, Lawrence Radiation Laboratory, Livermore, California (May 14, 1962).

[27] Cunningham, R. F., and P. F. Santarelli, "Computer Techniques for Parts Cataloging," *Proceedings of the 7th Annual Convention of the Society of Technical Writers and Editors,* Chicago, Illinois (April 1960).

[28] *Chemical-Biological Activities,* sample issue prepared by the Chemical Abstracts Service and published by the American Chemical Society (September 1962).

[29] Garfield, E., and I. H. Sher, "New Factors in the Evaluation of Scientific Literature Through Citation Indexing" *American Documentation,* Vol. 14, No. 3 (July 1963).

Garfield, E., "Generic Searching by Use of Rotated Formula Indexes," *J. Chemical Documentation,* Vol. 3, No. 2, pp. 97–103 (April 1963).

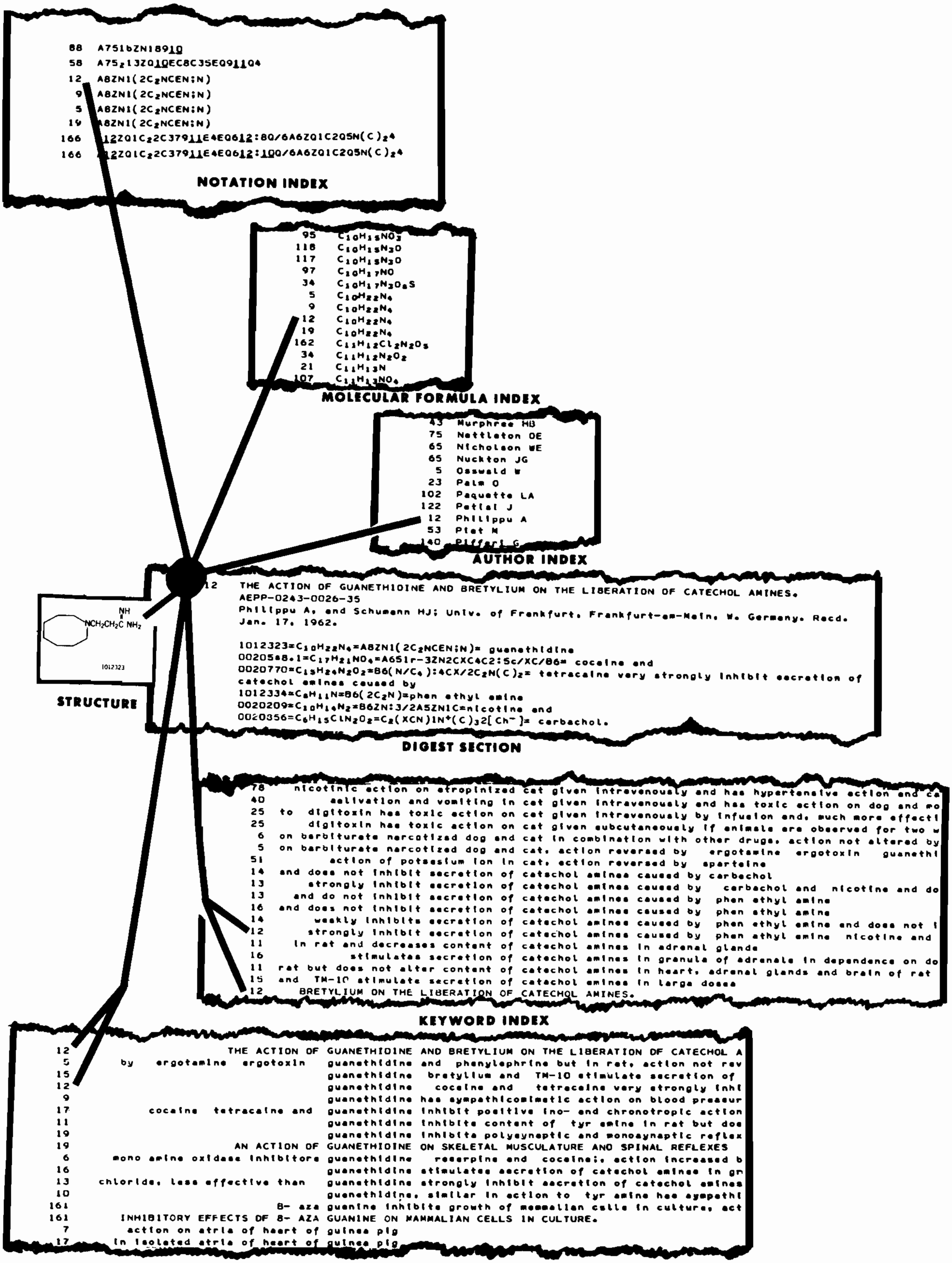


FIG. 7-14 Examples of computer-prepared indexes for *Chemical-Biological Activities.*

```
          BORN LK      56 AM DOC                      7  167
BORN M------------*51*RESTL UNIV----------□              161
          MORRIS JC    54 AM DOC                      5  117
      --------------51-RESTL UNIV----------□             162
          MORRIS JC    54 AM DOC                      5  117
      --------------51-RESTL UNIV----------□             282
          MORRIS JC    54 AM DOC                      5  117
BORNSTEIN HA-------*61*AM DOC--------------          12  254
          REES AM      62 AM DOC                     13   93
BOROS G------------*55*BOT WOERT-----------□
          OATFIELD H   58 AM DOC                      9  238
BOSE A-------------*56*MASS INST T REPT 309□
          ZADEH LA     61 IRE T INF T               IT7  139
      --------------59-IRE T CIRC THEORY--- CT6   30
          ZADEH LA     61 IRE T INF T               IT7  139
      --------------59-T 1959 INT SYMP CIRC□
          ZADEH LA     61 IRE T INF T               IT7  139
BOSE AG------------*  *MIT TECHN REP 309---□
          BRICK DB     62 IRE T INF T               IT8  S35
      --------------56-MIT RES LAB TR 309--□
          ORMSBY JF    61 J ACM                       8  440
      --------------56-TECHN REP 309 RES LA□
          KATZENEL. J  62 INF CONTR                   5  108
      --------------59-IRE T INF THEORY---- IT5   30
          BRICK DB     62 IRE T INF T               IT8  S35
          SWERLING P   61 IRE T INF T               IT7  131
BOSE RC------------*47*SANKHYA-------------           8  107
          BOSE RC      60 INF CONTR                   3   68
      --------------52-J AM STAT ASS--------         47  151
          NEUMANN PG   62 INF CONTR                   5   72
      --------------59-INF CONTR-----------           2  183
          GREY LD      62 IRE T INF T               IT8  200
          NEUMANN PG   62 INF CONTR                   5   72
      --------------60-INF CONTR-----------           3   68
          BANERJI RB   61 INF CONTR                   4    1
          BARTEE TC    62 IRE T INF T               IT8  S17
          BOSE RC      60 INF CONTR                   3  279
          CALINGAE. P  61 J ACM                       8  186
          ELIAS P      61 IRE T INF T               IT7  128
          GORENSTE. D  60 INF CONTR                   3  291
          MANN HB      62 INF CONTR                   5  153
          MEGGITT JE   61 IRE T INF T               IT7  234
          SELLERS FF   61 IRE T INF T               IT7  276
          STONE JJ     61 INF CONTR                   4  324
          ZETTERBE. L  62 IRE T INF T               IT8   13
      --------------60-INF CONTR-----------           3  279
          ELIAS P      61 IRE T INF T               IT7  128
          MANN HB      62 INF CONTR                   5  153
          STONE JJ     61 INF CONTR                   4  324
BOTTENBRUCH H------*61*ORNL REP 3148-------□
          BOTTENBR. H  62 J ACM                       9  161
BOULDING K---------*56*GEN SYST------------□              11
          SCHULTZ L    62 AM DOC                     13  288
      --------------56-IMAGE---------------□             166
          BUCHANAN B   58 AM DOC                      9  114
BOURGEOIS P--------*  *UNSPECIFIED---------□
          BONN GS      58 ICSI                        2 1441
      --------------32-TERM DOC------------□             101
          WAGNER FS    60 AM DOC                     11  102
BOURNE CP----------*58*MAG DAT-------------                6
          BONN GS      62 AM DOC                     13  301
      --------------61-INF CONTR-----------           4   48
          BOURNE CP    61 J ACM                       8  538
```

FIG. 7-15 Sample of a citation index prepared with computer equipment.

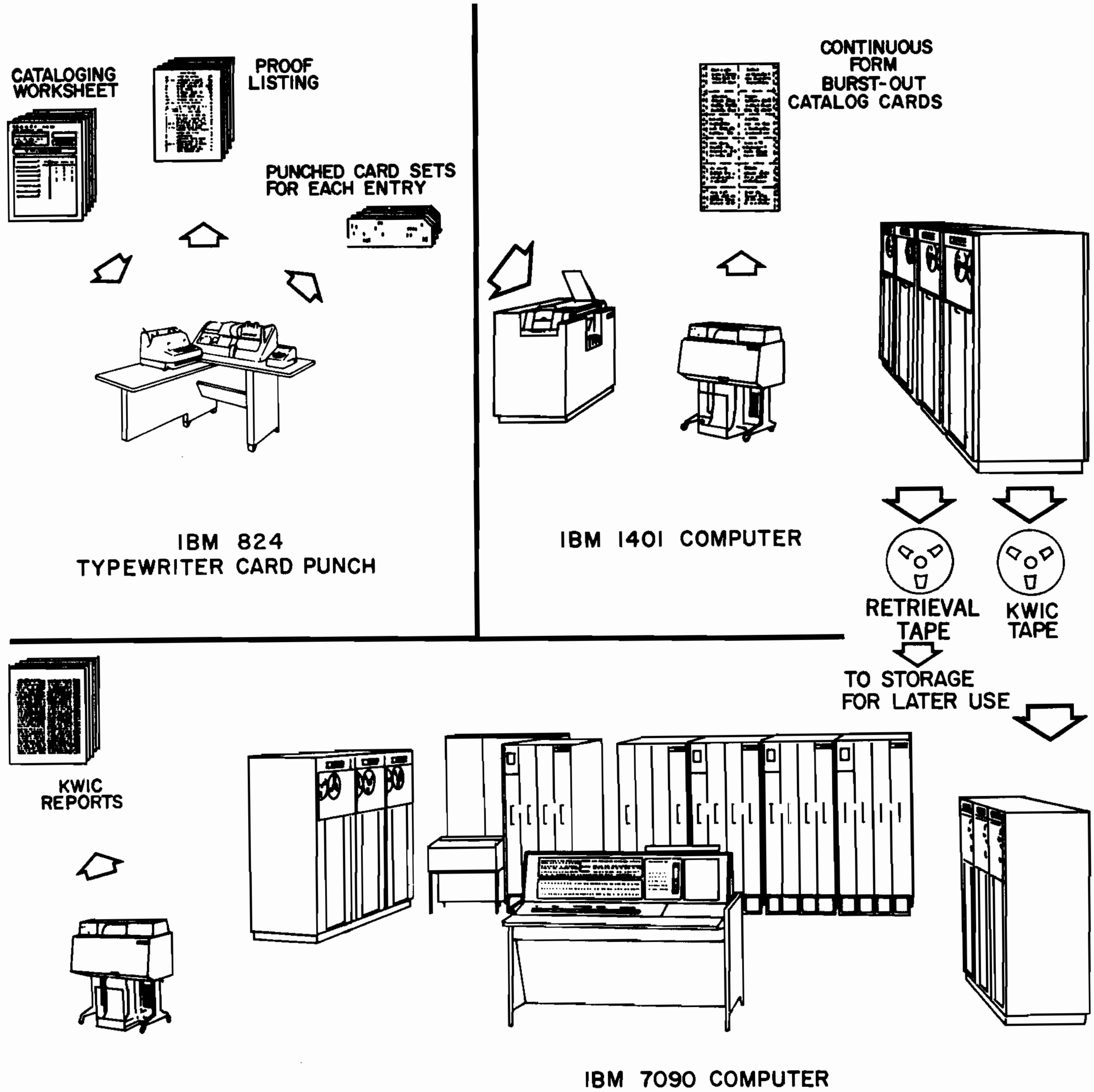


FIG. 7-16 Flow chart for a portion of Lockheed's Technical Information Center mechanization program.

Another example of the use of computers to prepare an index is the Tabledex Index, described previously in Chapt. 2.[80]

[80] Ledley, R. S., et al., *Selected Bibliography of the International Geophysical Year: An Example of Tabledex Formats,* National Biomedical Research Foundation Report No. 62071/18100 (1962).
Ledley, R. S., *Programming and Utilizing Digital Computers* (McGraw-Hill Book Co., New York, 1962), pp. 502–503, 505–506.

One particularly comprehensive computer indexing and cataloging system (shown in flow chart form in Fig. 7-16) is in operation at Lockheed Missile and Space Company in Palo Alto, California. This mechanized system for the reporting of incoming documents (approximately 2500 titles per month) produces a KWIC index of incoming technical reports, alphabetical author and source indexes for these reports, and from 60,000 to 75,000 catalog cards each month for filing at several locations. In addition, a magnetic tape is generated for file-searching programs, and a selective dissemination program is planned. Representative samples of the catalog cards printed by the computer in this system are shown in Fig. 7-17. The catalog cards are printed on pre-perforated stock that can easily be separated. Other organizations have also started preparing 3-by-5-inch catalog cards with computer equipment. The information for these catalog cards was initially keypunched and read into the computer, which then sorted and assembled the information to prepare and print the cards in proper filing sequence.

If multiple copies are required, then the printed copy from a computer's high-speed printer may be used directly as a reproduction copy for subsequent printing operations. However, the computer printers have four major drawbacks: the printing is all in capital letters, the characters are printed with a relatively few number of characters per line, the number of different characters available for printing is very limited, and only under very special circumstances can the printer print in more than one type font. One recent improvement is the IBM 1403 computer printer, which can be adapted to print both uppercase and lowercase letters (see examples in Fig. 7-14).

In the examples of the preceding paragraphs, brief descriptions were given of a few ways in which computers were used to assist in the preparation of subject indexes. However, in all these examples, the computer was used entirely for clerical tasks—it sorted and put things in order, it arranged different displays of material, and it printed information onto pages in a specified manner. In no case did the computer actually do any subject analysis of the material being indexed. Work of this type has been called "computer indexing" even though no real subject analysis was done by the computer and its program. It would be extremely useful if some subject analysis and indexing could be done by computers, and a considerable amount of research has been done to develop and study techniques for true "computer indexing."[31]

If a computer were to be used to perform some real subject analysis rather than some simple descriptive cataloging, what limitations would be imposed by the equipment, and what type of computer equipment would be necessary, rather than tab card equipment? The basic technique might depend on such things as a statistical analysis of the file item to determine the frequency with which certain words appear, or it might depend on the use of large tables of allowable indexing words that might occur in the text, or it might depend upon some linguistic and logical analysis of the text material. In any case the particular technique usually requires some relatively powerful (i.e., high-speed, large-memory) computer equipment.

[31] Edmundson, H. P., and R. E. Wyllys, "Automatic Abstracting and Indexing—Survey and Recommendations," *Communications of the Association for Computing Machinery,* Vol. 4, No. 5, pp. 226–234 (May 1961).

The basic computer capabilities that are exploited in such an analysis are: (1) the performance of some statistical analyses of the text material (word frequency counts, etc.); (2) the performance of some complex table look-up operations from large tables stored in the computer memory; and (3) the performance of logical operations (e.g., word comparison for correlation studies) between a large number of words.

The actual quality of the machine-produced index may depend upon the characteristics and the features of the particular computer used—some compromises may have to be made because of such things as the limited amount of memory available. Because of this, the type and quality of indexing may depend upon the computer chosen, although most of the medium and large computer systems would provide comparable performance for this task. The usual practice is to use whatever computer is handy. Whereas Chapt. 2 discussed indexing independent of any manner of implementation, we now see that the particular choice of equipment may have an influence on the product of the computer-prepared index.

To date, the different approaches to achieving computer-derived indexes by examining the text material in machine-readable form have been along the following general lines:

1. Extract and alphabetize or otherwise group words from the computer text that appear in a master tape of allowable subject headings.
2. Choose index entries on the basis of a statistical study of the frequency of word usage in the computer text.
3. Choose index entries on the basis of a statistical study of word frequency, location, and correlations.

Each of these general approaches is discussed in more detail in the following paragraphs.

STL-GM-TR-0000-00069 UN
SPACE TECHNOLOGY LABS.
EXPERIMENTAL INVESTIGATION OF THE
INFLUENCE OF MECHANICAL BAFFLES ON THE
FUNDAMENTAL SLOSHING MODE OF WATER IN A
CYLINDRICAL TANK.
HOWELL, E. + EHLER, F.G.
LOS ANGELES, CALIF. 6 JULY 56.
43P. C-1-P

1.T. 2.LIQUIDS-OSCILLATION 2A.DAMPING--
DETERMINATION 2B.FUEL TANKS--VIBRATION
3A.HOWELL, E. 3B.EHLER, F.G. 4A.STL-GM-
TR-69 4B.STL-GM-TR-0000-00069
4C.STL-GM-45.3-87

STL-GM-TR-0000-00069 UN
SPACE TECHNOLOGY LABS.
EXPERIMENTAL INVESTIGATION OF THE
INFLUENCE OF MECHANICAL BAFFLES ON THE
FUNDAMENTAL SLOSHING MODE OF WATER IN A
CYLINDRICAL TANK.
HOWELL, E. + EHLER, F.G.
LOS ANGELES, CALIF. 6 JULY 56.
43P. C-1-P

1.T. 2.LIQUIDS-OSCILLATION 2A.DAMPING--
DETERMINATION 2B.FUEL TANKS--VIBRATION
3A.HOWELL, E. 3B.EHLER, F.G. 4A.STL-GM-
TR-69 4B.STL-GM-TR-0000-00069
4C.STL-GM-45.3-87

STL-GM-59-0000-14217 S
SPACE TECHNOLOGY LABS.
INERTIAL GUIDANCE SYSTEMS. STATE OF THE
ART. A LITERATURE SURVEY OF CLASSIFIED
REPORTS.
ANDREWS, K.B. + SLOANE, M.N.
LOS ANGELES, CALIF. 30 DEC. 59.
114P. C-1-P

1.T. 2A.BIBLIOGRAPHY(INERTIAL GUIDANCE
SYSTEMS) 2B.INERTIAL GUIDANCE SYSTEMS--
BIBLIOGRAPHY 3A.ANDREWS, K.B. 3B.SLOANE,
M.N. 4A.STL-GM-59-0000-14217 4B.STL-B-28
4C.L-128-561

STL-GM-59-0000-14217 S
SPACE TECHNOLOGY LABS.
INERTIAL GUIDANCE SYSTEMS. STATE OF THE
ART. A LITERATURE SURVEY OF CLASSIFIED
REPORTS.
ANDREWS, K.B. + SLOANE, M.N.
LOS ANGELES, CALIF. 30 DEC. 59.
114P. C-1-P

1.T. 2A.BIBLIOGRAPHY(INERTIAL GUIDANCE
SYSTEMS) 2B.INERTIAL GUIDANCE SYSTEMS--
BIBLIOGRAPHY 3A.ANDREWS, K.B. 3B.SLOANE,
M.N. 4A.STL-GM-59-0000-14217 4B.STL-B-28
4C.L-128-561

AFBMD-TR-59-16 (IV) SRD
SPACE TECHNOLOGY LABS.
SURVEY OF FUTURE REQUIREMENTS AND
CAPABILITIES OF BALLISTIC MISSILE AND SPACE
VEHICLE POWER SYSTEMS. PART IV SUMMARY AND
CONCLUSIONS.
SOHN, R.L.
LOS ANGELES, CALIF. 23 NOV. 59.
62 P. C-TO-S 50

1.T. 2A.SPACESHIPS--POWER SUPPLIES
2B.BALLISTIC MISSILES--POWER SUPPLIES
2C.POWER SUPPLIES--ANALYSIS 2D.SOLAR ENERGY
--STORAGE 2E.PHOTOELECTRIC CELLS 3A.SOHN,
R.L. 4A.L-127-515 4B.AFBMD-TR-59-16(IV)
4C.STL-TR-59-0000-00831-PT.IV
5A.AF-04(647)-309

AFBMD-TR-59-16 (IV) SRD
SPACE TECHNOLOGY LABS.
SURVEY OF FUTURE REQUIREMENTS AND
CAPABILITIES OF BALLISTIC MISSILE AND SPACE
VEHICLE POWER SYSTEMS. PART IV SUMMARY AND
CONCLUSIONS.
SOHN, R.L.
LOS ANGELES, CALIF. 23 NOV. 59.
62 P. C-TO-S 50

1.T. 2A.SPACESHIPS--POWER SUPPLIES
2B.BALLISTIC MISSILES--POWER SUPPLIES
2C.POWER SUPPLIES--ANALYSIS 2D.SOLAR ENERGY
--STORAGE 2E.PHOTOELECTRIC CELLS 3A.SOHN,
R.L. 4A.L-127-515 4B.AFBMD-TR-59-16(IV)
4C.STL-TR-59-0000-00831-PT.IV
5A.AF-04(647)-309

SRI-2532-IR-7 S
STANFORD RESEARCH INST.
AIRBORNE JAMMING ANTENNA STUDY STUDY
OF VOLTAGE BREAKDOWN, VARIABLE-GAIN
ANTENNAS,MULTIDIMENSIONAL PATTERN
SYNTHESIS, ANTENNAS FOR DECEPTION JAMMING...
SCHARFMAN, W.E.,ET AL
MENLO PARK, CALIF. MAR. 61.
127 P. C-1-P 68

1.T. 2A.RADAR JAMMING SYSTEMS(AIRBORNE)
2B.ANTENNAS--OPERATION 2C.ANTENNA
RADIATION PATTERNS 3A.SCHARFMAN, W.E.,ET AL
4A.L-129-556 4B.SRI-2532-IR-7
5A.AF-33(616)-5584

SRI-2532-IR-7 S
STANFORD RESEARCH INST.
AIRBORNE JAMMING ANTENNA STUDY STUDY
OF VOLTAGE BREAKDOWN, VARIABLE-GAIN
ANTENNAS,MULTIDIMENSIONAL PATTERN
SYNTHESIS, ANTENNAS FOR DECEPTION JAMMING...
SCHARFMAN, W.E.,ET AL
MENLO PARK, CALIF. MAR. 61.
127 P. C-1-P 68

1.T. 2A.RADAR JAMMING SYSTEMS(AIRBORNE)
2B.ANTENNAS--OPERATION 2C.ANTENNA
RADIATION PATTERNS 3A.SCHARFMAN, W.E.,ET AL
4A.L-129-556 4B.SRI-2532-IR-7
5A.AF-33(616)-5584

RADC-TN-60-192 UN
SPERRY-GYROSCOPE CO.
METAL-TO-CERAMIC SEAL TECHNOLOGY STUDY.
COLE, S.S.,JR.,ET AL
ELECTRONIC TUBE DIV.,GREAT NECK, N.Y.
31 MAR. 60 TO 30 JUNE 60.
48 P. C-1-P

1.T. 2A.SEALS--MATERIALS 2B.METAL SEALS
2C.CERAMIC MATERIALS--APPLICATIONS
3A.COLE,S.S.,JR.,ET AL 4A.SPERRY-NA-
8240-8184-4 4B.AD-246-644 4C.RADC-TN-
60-192 5A.AF-30(602)-2047

RADC-TN-60-192 UN
SPERRY-GYROSCOPE CO.
METAL-TO-CERAMIC SEAL TECHNOLOGY STUDY.
COLE, S.S.,JR.,ET AL
ELECTRONIC TUBE DIV.,GREAT NECK, N.Y.
31 MAR. 60 TO 30 JUNE 60.
48 P. C-1-P

1.T. 2A.SEALS--MATERIALS 2B.METAL SEALS
2C.CERAMIC MATERIALS--APPLICATIONS
3A.COLE,S.S.,JR.,ET AL 4A.SPERRY-NA-
8240-8184-4 4B.AD-246-644 4C.RADC-TN-
60-192 5A.AF-30(602)-2047

FIG. 7-17 Sample catalog cards prepared by computer equipment.

One of the easiest ways, conceptually, to mechanize the indexing operation is to check each word of the text against a previously compiled list or thesaurus of possible subject words. The text words that coincide with the table entries can then be extracted and used as indexing terms. This technique has been used by Doyle.[32] It had also been suggested earlier by Luhn, who used a table entry for the additional function of providing generic or notional indexing terms to replace the original word of the text.[33] For a fixed collection of file items, this table look-up operation may possibly provide satisfactory results. However, the method is somewhat inflexible, and cannot easily handle collections of items that do not have the specific or anticipated indexing words embedded in the text, or subject fields that do not have a well-developed vocabulary, or collections that are multi-disciplinary. If a simple table were used, the indexing product would probably be a list of specific words from the text, similar to the results achieved by other word-oriented indexing schemes such as the Uniterm and key word systems. If a thesaurus or more elaborate table were used, then the indexing product might be extended to provide larger notional or generic groups for more general subject indexing.

One experimental study by Montgomery and Swanson has indicated that for a large proportion of the entries in selected portions of a classified bibliography, the *Index Medicus,* a large fraction of the titles contained words which were synonymous with or identical to the words of the corresponding subject heading. It was suggested that given a list of titles, the *Index Medicus* subject heading list, and a synonym list or thesaurus group for each subject heading, a computer could be programed to determine which subject heading should be assigned to each article, and that about 86 percent of such assignments would be the same as assignments made by human indexing.[34] This still leaves 14 percent to be assigned by human indexers, and thus precludes the possibility of having a completely automatic indexing technique.

Luhn's early work has received the most attention of all of the automatic indexing and abstracting techniques and relies heavily on statistics of frequency of word usage.[35] The indexing method makes two basic assumptions: (1) the more often a word appears in a document, the more likely it is that the word is representative of the subject matter of that document; and (2) depending on their positions within a sentence, the representative words can be considered to be related. The indexing procedure consists of counting the number of times that all the words are used in the text and (with the exception of common or junk words) selecting the most "predominant" words, and then looking at the relative sentence position of the predominant and less predominant words to arrive at word patterns which are unique to that document. It is assumed that similar word patterns for two documents suggest similar contents. After the word counts have been completed, the remaining words or word roots are sorted according to the number of times they are used in the text. The words which are used most often in the document are selected for further analysis. That is, the 10, 20, or n most-used words are extracted from the word list. The procedure could be terminated at this point with a list of the words that were used most often in the document; and this list of terms might look very much like the results achieved by word-oriented indexing schemes. Some editing would probably be required to handle synonyms and special cases, but the majority of the indexing effort (i.e., a "first draft" index) would already have been done with very little human intervention. Luhn then suggests the use of a thesaurus in order to categorize the various words which were selected. This would have the effect of providing some post-editing since it would probably handle some of the synonym problems, and would point out the words which had no previous entry in the thesaurus. Luhn's next step is to examine the position of each index word and its proximity to other significant words in the sentence in order to determine major and minor notions. The final result of the indexing operation is then a serial notation which describes the various notions that

[32] Doyle, L., "Programmed Interpretation of Text as a Basis for Information Retrieval Systems," *Proceedings of the 1959 Western Joint Computer Conference*, pp. 60–63 (Institute of Radio Engineers, New York, 1959).

[33] Luhn, H. P., "A Statistical Approach to Mechanized Encoding and Searching of Literary Information," *IBM Journal*, Vol. 1, No. 4, pp. 309–317 (October 1957).

[34] Montgomery, C., and D. R. Swanson, "Machinelike Indexing by People," *American Documentation*, Vol. 13, No. 4, pp. 359–366 (October 1962).

[35] Luhn, H. P., *Potentialities of Auto-Encoding of Scientific Literature*, Report RC-101, IBM Research Center, Yorktown Heights, New York (May 1959).

were suggested by the document, as well as the pattern of linkage between the various notions.[36]

Baxendale's method of indexing is similar in many respects to some of Luhn's techniques, and also makes the basic assumption that the more often a word appears in a document, the more it becomes representative of the contents of the document.[37] However, Baxendale obtains her word frequency distributions from selected portions of the text instead of using the entire text, and with the exception of developing groups of coordinated index words, her indexing procedure is not carried beyond the point of obtaining a list of the most-used words. Three different techniques were suggested for selecting portions of the text in order to reduce the volume of text which is to be analyzed, and all take advantage of certain grammatical or syntactical characteristics of the English language. They may be characterized as:

1. Exploitation of topic sentences
2. Deletion of non-discriminating index words
3. Selection of prepositional phrases

The first technique takes advantage of an indexing or abstracting feature built in by the author of the document. In general, students have traditionally been taught that, for technical writing, the best emphasis can be achieved if they put the prime thought of a paragraph in either the first or last sentence. Preliminary investigations of sample text material indicated that, 85 percent of the time, the topic sentence actually appeared as the first sentence; and 7 percent of the time, the topic sentence appeared as the last sentence. The exploitation of this particular writing habit provides a technique by which some very essential parts of a document, possibly a good distillation of the document, may be selected on a semi-automatic basis. This technique is not appropriate for documents with unusual styles or formats, and would probably overlook descriptive terms (such as names or proper nouns) appearing in places other than the topic sentences. The second technique is used to remove the common or non-descriptive words, and experiments showed that on the average, over 50 percent of the text material could be removed by this screening operation. The screening process used a table of approximately 150 words (including all pronouns, articles, conjunctions, conjunctive adverbs, copulas, and auxiliary verbs, as well as quantitative adjectives) which had very little discriminating power for indexing purposes. Word frequency distributions were developed from the words that remained after screening. The third technique considers that in addition to the topic sentence, there are other meaningful syntactical units within the text. In particular, it is argued that prepositional phrases may be the most central features of the text, and that the phrase is likely to reflect the content of an article more closely than any other simple construction. If the selection of the phrases is to be made automatically, then some syntactic characteristic must be found which will permit this. In the case of the prepositional phrase, the preposition itself can be used as the indicator to initiate the selection of the phrases. There are less than 50 common prepositions, so that a simple table look-up operation on the words of the text will uncover the starting points of the phrases. This extraction of phrases also reduces the volume of words which would subsequently be used to construct a word frequency distribution. This third method has the added advantage of easily producing lists of coordinated index terms (*bound descriptors*) and some key words in context as they appeared in the phrases. These three methods all provide a list of index terms as the final product. Presumably, any of these techniques could be extended to do such things as check the selected index words against a thesaurus or dictionary in order to consolidate words with similar root stems, or consolidate synonyms, or redescribe the words as part of some generic family.

Another possible method for automatically selecting index words has been suggested by Doyle, which utilizes some probability features, and requires a statistical analysis of the text material in the total collection.[38] Doyle's basic assumption is that a

[36] Luhn, H. P., Auto-Encoding of Documents for Information Retrieval Systems," in *Modern Trends in Documentation,* M. Boaz, editor, pp. 45–58 (Pergamon Press, New York, 1959).

Luhn, H. P., *A Serial Notation for Describing the Topology of Multidimensional Branched Structures* (*Nodal Index for Branched Structures*), Report RC-27, IBM Research Center, Yorktown Heights, New York (December 1955).

[37] Baxendale, P. B., "Machine-Made Index for Technical Literature—An Experiment," *IBM Journal of Research and Development,* Vol. 2, No. 4, pp. 354–361 (October 1958).

Baxendale, P. B., "An Empirical Model for Machine Indexing," in *Machine Indexing: Progress and Problems,* pp. 207–218 (American University, Washington, D.C., 1961).

[38] Doyle, L., fn. 32.

good subject word should have certain other words which coexist with it, with a frequency much greater than pure chance. He also suggests that the natural characterization and organization of information can come from analysis of frequencies and distributions of words in libraries.[39] In order to test or use this hypothesis, a statistical study must be made to determine the probabilities of occurrence of each of the known indexing words. These probabilities will be influenced by the habitual word patterns of the individual authors, and by the number of different disciplines which are represented by the collections. Furthermore, automatic indexing cannot begin until a large collection of documents has been assembled and statistically analyzed. To date, this method or hypothesis has not been rigorously tested, and has only been suggested as an avenue for further study.

As a result of one empirical study, Vickery has suggested that a frequency count of words occurring in a text can, in some cases, and to a first approximation, select words significant for retrieval, provided that an extensive list of common or junk words is excluded.[40] He notes that in the computer-derived KWIC index to *Chemical Titles,* at least 960 common words are excluded. Some preliminary tests have been made by Maron, who suggests that statistics on kind, frequency, location, and order of selected clue words are adequate to make reasonably good predictions about the subject matter of documents containing those words.[41] Thus, on the basis of the occurrence of selected clue words, a computer might be able to decide to which of many subject categories the document in question belongs.

Borko has demonstrated an empirical technique by which a crude classification of material can be derived semi-automatically from a study of the text of abstracts of documents.[42] From a list of significant words that occurred in the text material 20 or more times, 90 words were selected for use as index terms. A correlation study was made to determine relationship of each of these index terms to every other index term. A factor analysis and study of this correlation matrix yielded 10 meaningful classification categories. After Borko assigned the labels, these experimentally derived categories were shown to be compatible but not identical to one of the standard classification schemes already in use for this body of material.

As a concluding comment on this question of computer-derived indexes, it should be noted that there are some indications based on empirical studies of the relationship between word frequency and word rank (Zipf's law), that the frequency of occurrence of single words or word pairs does not provide enough distinct steps for adequate discrimination.[43] This same lack of discrimination exists in measures of word lengths, even though the words are generally more specific and informative as the word length is increased. A syntactical approach may be useful here to augment the machine indexing or abstracting approaches that utilize semantic or statistical techniques. Initial studies have indicated a strong correlation between the sentence structure and the information represented by the sentence.[44]

Data processing equipment can select certain words from the text which satisfy certain statistical properties. However, it is unlikely that in the near future, any machine indexing will approach the quality obtained by manual indexing. We can look forward to seeing continuing empirical studies being made along these lines as it becomes easier to obtain large volumes of text material in machine-readable form. The difficulty of obtaining large volumes of text material in machine language may eventually

[39] Doyle, L. B., "Semantic Road Maps for Literature Searchers," *Journal of the Association for Computing Machinery,* Vol. 8, No. 4, pp. 553–578 (October 1961).

Doyle, L. B., "Indexing and Abstracting by Association," *American Documentation,* Vol. 13, No. 4, pp. 378–390 (October 1962).

[40] Vickery, B. C., "Statistical Methods in Indexing," *Revue de la Documentation,* Vol. 28, No. 2, pp. 56–62 (May 1961).

[41] Maron, M. E., "Automatic Indexing: An Experimental Inquiry," *Journal of the Association for Computing Machinery,* Vol. 8, No. 3, pp. 404–417 (July 1961).

[42] Borko, H., "The Construction of an Empirically Based Mathematically Derived Classification System," *Proceedings of the 1962 Spring Joint Computer Conference,* pp. 279–289 (The Institute of Radio Engineers, New York, 1962).

[43] Ohlman, H., "Historical Development of Mechanical Indexing," paper presented at the Annual Convention of the American Documentation Institute, Berkeley, California, October 1960.

Zipf, G. K., *The Psychobiology of Language* (Houghton Mifflin Co., Boston, Massachusetts, 1935).

Zipf, G. K., *Human Behavior and the Principal of Least Effort* (Addison-Wesley Publishing Co., Cambridge, Massachusetts, 1949).

[44] Climenson, W. D., N. H. Hardwick, and S. N. Jacobson, "Automatic Syntax Analysis in Machine Indexing and Abstracting," *American Documentation,* Vol. 12, No. 3, pp. 178–183 (July 1961).

be reduced somewhat by the availability of automatic page-reading equipment.[45]

Abstracting. Most of the research work reported to date on abstracting by computer, has followed the initial lead of Luhn.[46] The basic approach suggested for computer abstracting is very similar to the techniques Luhn proposed for computer indexing. The assumption that the frequency with which a word is used furnishes a useful measure of the importance of that word is basic to this abstracting technique as it was to the indexing technique. It is also assumed that the relative position within a sentence of words with a given level of significance furnishes a measure of importance of the sentence. The significance factor of a sentence is based on a combination of these two measures. The text material is scanned to obtain the word frequency distributions. As in the indexing technique, the common words are screened out and are not included in the frequency distributions. Also, the words with similar root stems are consolidated together as one entry.

The program then essentially looks for the sentence in which the most frequently occurring words were found closest to each other. For each sentence in the text, a number is computed which reflects the number of occurrences of significant words within that sentence as well as their closeness together within the sentence. All the sentences are ranked according to this significant factor, and a chosen number of the highest-ranking sentences are selected to be assembled together as the abstract. Thus the abstract actually consists of extracts of the article, and consists only of the author's original words, with no combination or summarization.

A modification to this basic auto-indexing, auto-abstracting, or auto-condensation method has been proposed by Bar-Hillel.[47] He suggests that instead of ranking the selected words according to their frequency of occurrence within the text, they should be ranked according to the difference between their actual frequency in the document and the average frequency with which they might occur in the entire library of documents. And as with the earlier schemes, a chosen segment of this ranking would be used for indexing or abstracting purposes. This suggestion has received additional attention by Edmundson and Wyllys.[48]

There are strong arguments that an abstract should be a re-write instead of an extraction of selected sentences from the text. And in contrast to the critics who reject the mechanization of indexing and abstracting as impractical for the foreseeable future, there are also great numbers of people who are convinced that such techniques must be used and improved, regardless of their limitations, in order to supplement the increasing amount of intellectual effort which is currently required and will be required in the future. Several studies have been conducted recently to determine what constitutes a good abstract and what specific instructions might be followed in order to produce an acceptable abstract.[49] Preliminary studies indicate that the current cost of obtaining an abstract by computer techniques is on the order of 48 dollars per abstract for a 4400-word document, which is considerably higher than the cost of doing it manually.[50]

[45] Cornelius, M. E., "Machine Input Problems for Machine Indexing: Alternatives and Practicalities," pp. 41–49 in *Machine Indexing: Progress and Problems* (American University, Washington, D.C., 1961).

[46] Luhn, H. P., "The Automatic Creation of Literature Abstracts," *Proceedings of the 1958 Institute of Radio Engineers National Convention,* New York. Reprinted in the *IBM Journal,* Vol. 2, No. 2, pp. 159–165 (April 1958).

Luhn, H. P. *An Experiment in Auto-Abstracting,* an un-numbered report of the IBM Research Center, Yorktown Heights, New York (November 1958).

Savage, T. R., *The Preparation of Auto-Abstracts on the IBM 704 Data Processing System,* an un-numbered report of the IBM Research Center, Yorktown Heights, New York (November 1958).

Resnick, A., and T. R. Savage, *A Re-Evaluation of Machine Generated Abstracts,* Research Report RC-230, IBM Research Center, Yorktown Heights, New York (March 1, 1960).

Rath, G. J., A. Resnick, and T. R. Savage, "Comparisons of Four Types of Lexical Indicators of Content," *American Documentation,* Vol. 12, No. 2, pp. 126–130 (April 1961).

Rath, G. J., A. Resnick, and T. R. Savage, "The Formation of Abstracts by the Selection of Sentences," *American Documentation,* Vol. 12, No. 2, pp. 139–143 (April 1961).

[47] Bar-Hillel, Y., "The Mechanization of Literature Searching," *Proceedings of a Symposium on the Mechanization of Thought Processes,* held at the National Physical Laboratory, Teddington, Middlesex, England, in November 1958 (Her Majesty's Stationery Office, 1959).

[48] Edmundson, H. P., and R. E. Wyllys, fn. 31.

[49] De Lucia, A., "Index-Abstract Evaluation and Design," paper presented at the annual convention of the American Documentation Institute, December 13, 1962, Hollywood-by-the-Sea, Florida.

Borko, H., *Criteria for Acceptable Abstracts: A Survey of Abstractors' Instructions,* Report TM-759 of the System Development Corp., Santa Monica, California (November 1, 1962).

[50] Hensley, C. B., "A Note on Changing Economics of Auto-Abstracting," paper presented at the annual convention of the American Documentation Institute, December 14, 1962, Hollywood-by-the-Sea, Florida.

File Searching

The earliest use of computers for searching an index to a collection of technical publications appears to have been made in 1954.[51] A coordinate indexing system was used for a collection of about 25,000 documents, each of which was indexed by about 8 descriptors from a dictionary of about 9600 descriptors. At that time, about 16 searches per week were performed using the computer 3 times per week. A total of about 12 minutes of machine time were required for posting all the new entries to the file and performing the 16 searches. The entire index was stored on approximately one-third of a reel of magnetic tape. As many as 75 searches could be conducted simultaneously. The result of a search was a list of combinations of search terms and associated document numbers that satisfied the search criteria. This pioneer installation was the first to experiment with such a system, and quickly demonstrated some shortcomings, such as the fact that it might not be able to provide the answers in as short a time as the manual system, it was expensive, and it did not economically permit browsing by the user.[52]

A recent review of this system indicated that after 8 years of operation, a coordinate indexing system was still being used, but now on an IBM 7090 and with a collection of 90,000 reports.[53] The intellectual problems of subject indexing were still there; and maybe amplified—but some clerical operations were minimized. It was felt that the new system is providing better service than the old machine system did, although it is still more convenient in many cases to use the manual system. The main problem with the new machine system is the fact that the user cannot browse through the index, and the degree of success of searching is unknown until the computer search has been finished. The search program is run once each week, and the search requests for the week are all batched together for this one run, since many users are willing to wait up to a week to get their search results. The actual cost to the user depends upon the amount of computer time used and the number of other users in the batch who will share the cost of that run. Since the actual machine time depends upon several factors such as the file size and the number of items to be printed, it is very difficult to establish any fixed estimate of the cost of running this program. However, current experience shows, for example, that a search with 21 questions requires approximately $250 worth of machine time. The costs seem to range from 8 to 25 cents per bibliographic file item produced by the search. The initial programing and file preparation costs for this new system was in excess of $20,000.

At about this same time, a similar line of work was being pursued in the development of programs for the IBM 701 and IBM 704 to retrieve codes representing chemical structures from a large file of encoded chemical structures.[54] These programs searched the files at a rate of 10,000 compounds per minute per search question and provided chemical names as the search product. Computer output equipment is currently available that will permit the direct display and printout of representations of the chemical structures, in addition to the chemical names.[55] These programs were subsequently incorporated as part of a larger program for the IBM 704 to find undiscovered uses for commercially available chemicals.[56]

[51] Bracken, R. H., and H. E. Tillitt, "Information Searching with the 701 Calculator," *Journal of the Association for Computing Machinery,* Vol. 4, No. 2, pp. 131–136 (April 1957).

Tillitt, H., "An Application of an Electronic Computer to Information Retrieval," pp. 67–69 in *Modern Trends in Documentation,* M. Boaz, editor (Pergamon Press, New York, 1959).

[52] Bloomfield, M., "Evaluation of Coordinate Indexing at the Naval Ordnance Test Station," *American Documentation,* Vol. 8, No. 1, pp. 22–25 (January 1957).

[53] Kruse, C. J., "The Use of Electronic Computers for Information Retrieval at the Naval Ordnance Test Station," paper presented at the annual convention of the Special Libraries Association, Washington, D.C., May 1962.

[54] Opler, A., and T. R. Norton, "New Speed to Structural Searches," *Chemical and Engineering News,* Vol. 34, No. 23, pp. 2812–2816 (June 4, 1956).

Opler, A., "Dow Refines Structural Searching," *Chemical and Engineering News,* Vol. 35, No. 33, pp. 92–96 (August 19, 1957).

[55] Opler, A., and N. Baird, "Display of Chemical Structural Formulas as Digital Computer Output," *American Documentation,* Vol. 10, No. 1, pp. 59–63 (January 1959).

Waldo, W. H., and M. DeBacker, "Printing Chemical Structures Electronically: Encoded Compound Searched Generically with IBM-702," *Proceedings of the International Conference on Scientific Information,* Vol. 1, pp. 711–730 (National Academy of Sciences, Washington, D.C., 1959).

Waldo, W. H., R. S. Gordon, and J. D. Porter, "Routine Report Writing by Computer," *American Documentation,* Vol. 9, No. 1, pp. 28–31 (January 1958).

[56] Findley, L. D., et al., *A Card Controlled Routine for Searching Chemical Compound Data with an IBM 704,* an

Prime-number coding, described earlier, in Chapt. 3, was used for coding a collection of organic chemicals for searching with the Pegasus computer. A collection of some 25,000 organic chemicals, together with a file of relevant physical, chemical, biological, and other properties were indexed by approximately 250 different structural features (e.g., ring systems, amines, chlorides). The first 208 prime numbers were allocated to 208 of the structural features, with the smaller primes allocated to the more frequently occurring features. Since a given chemical could be indexed by as many as 17 different features, the largest compound indexing number that could possibly occur would be something less than the 208th prime used 17 times, which is about 6×10^{52}. However, the largest compound number found in practice was about 10^{22}, considerably less than the maximum possible. In this particular case, prime number coding provided some definite advantages by reducing the amount of storage required and speeding up the search.[57]

The U.S. Patent Office has had a considerable amount of interest in chemical compound searching because of the potential value in its application to patent searching. Members of the Patent Office and the National Bureau of Standards have worked together for several years on experimental programs for chemical compound searching,[58] and the National Bureau of Standards has recently published a survey of computer programs that have been developed for this purpose.[59] Studies on the application of computers to chemical compound searching and the processing or manipulation of chemical structures have been pursued by a number of other organizations.[60]

Computers have been used in several instances to search files of equipment reliability or performance data. Also, one large aircraft company has put the detailed personnel records of its professional and technical staff on computer magnetic tape in order to search for individuals with particular skills. The detailed data for each employee represent a profile and cataloging of his skills and interests. The file includes details of work experience, educational background, special achievements, professional affiliations, and technical specialties.

As mentioned earlier, the first real application of computers to literature searching began in 1954.

un-numbered report from the Midwest Research Institute, Kansas City, Missouri (November 1958).

Opler, A., "Utilization of Computers for Information Retrieval," *Proceedings of the Fifth Annual Computer Applications Symposium* (Armour Research Foundation, Chicago, Illinois, 1958).

57 Cockayne, A. H., and E. Hyde, "Prime Number Coding for Information Retrieval," *The Computer Journal* (British), Vol. 3, No. 1, pp. 21–22 (April 1960).

58 Koller, H. R., E. Marden, and H. Pfeffer, "The HAYSTAQ System: Past, Present, and Future," *Proceedings of the International Conference on Scientific Information,* Vol. 2, pp. 1143–1179 (National Academy of Sciences, Washington, D.C., 1959).

Pfeffer, H., H. R. Koller, and E. Marden, "A First Approach to Patent Searching Procedures on Standard's Electronic Automatic Computer (SEAC)," *American Documentation,* Vol. 10, No. 1, pp. 20–26 (January 1959).

Ray, L. C., and R. A. Kirsch, "Finding Chemical Records by Digital Computers," *Science,* Vol. 126, No. 3278, pp. 814–819 (October 25, 1957).

Andrews, D. D., "Storing Numerical Values in an Inverted File for Mechanized Information Retrieval," *Journal of the Franklin Institute,* Vol. 270, No. 1, pp. 34–41 (July 1960).

59 Marden, E. C., and H. R. Koller, *Survey of Computer Programs for Chemical Information Searching,* National Bureau of Standards Technical Note 85 (February 1961), PB-161 586.

60 Opler, A., "A Topological Application of Computing Machines," *Proceedings of the 1956 Western Joint Computer Conference,* Institute of Radio Engineers, New York.

Opler, A., and N. Baird, "On the Automatic Manipulation of Representations of Chemical Structures," *American Documentation,* Vol. 10, No. 2, pp. 130–134 (April 1959).

Stokolova, N. A., "A System of Recording and Searching Classes of Chemical Compounds," *Foreign Developments in Machine Translation and Information Processing,* No. 21, JPRS: 8479, distributed by Office of Technical Services, U.S. Department of Commerce, Washington, D.C. (June 22, 1961).

Seyfer, A. L., and V. S. Shteyn, "Information Problems in the Field of Physical-Chemical Analysis and Their Solution by Means of Electronic Digital Computers," *Foreign Developments in Machine Translation and Information Processing,* No. 31, JPRS: 8479, distributed by Office of Technical Services, U. S. Department of Commerce, Washington, D.C. (June 22, 1961).

Lakhuti, D. G., and N. A. Stokolova, "The Problem of Searching Chemical Abstracts by Title," *Foreign Developments in Machine Translation and Information Processing,* No. 31, JPRS: 8479, distributed by Office of Technical Services, U.S. Department of Commerce, Washington, D.C. (June 22, 1961).

Anon., "The 'Filter' System for Searching Chemical Compounds According to Structural Characteristics," translation of a Russian-language brochure. English translation in *Foreign Developments in Machine Translation and Information Processing: USSR,* No. 41, JPRS: 8991, available from Office of Technical Services, U.S. Department of Commerce, Washington, D.C. (October 5, 1961).

Waldo, W. H., "Searching Two-Dimensional Structures by Computer," *Journal of Chemical Documentation,* Vol. 2, No. 1, pp. 1–2 (January 1962).

```
TO  FPD TECHNICAL INFORMATION CENTER
    BUILDING 100

                     D F BERG
                     K102, X2215
                     06/21/62

                     SEARCH 0000BF

           KEY WORDS SEARCHED

                    OPTIMUM
                    PERFORMANCE

 PLEASE SEND ON LOAN THE MATERIAL INDICATED BELOW.     TOTAL NO.   27
```

002362	003387	005592	008478	025080	026401
029487	035650	036486	039832	042707	044594
045109	045603	046614	048075	050628	052784
058395	058596	058700	059767	062637	063362
063477	065546	067284			

FIG. 7-18 Sample computer search printout showing selected document numbers.

Since that time, a large number of organizations have established and operated their own literature systems using a wide variety of computers. The IBM Research Laboratory in San Jose, California, established an experimental searching system in 1958, first with an IBM 305 RAMAC data processing system, and then later with an IBM 704 for a completely operational system. The RAMAC system stored an index to 5000 documents, and had the capacity to store an index to 25,000 documents. A normal search of the index produced a result in less than one minute.[61]

[61] Firth, F. E., "An Experiment in Literature Searching with the IBM 305 RAMAC," *Proceedings of the 1958 Western Joint Computer Conference,* Los Angeles (The Institute of Radio Engineers, New York).

Nolan, J. J., "Information Storage and Retrieval Using a Large Scale Random Access Memory," *American Documentation,* Vol. 10, No. 1, pp. 27–35 (January 1959).

In late 1958 the Flight Propulsion Division of the General Electric Company implemented a coordinate index searching system with an IBM 704. In 1961 the system was transferred to an IBM 7090 such as the one shown in Fig. 7-9. The coordinate index for over 65,000 documents is maintained on magnetic tape and searched at periodic intervals or upon demand. New file entries are posted at the same time that the searches are performed. The present program will accommodate up to 1300 searches at the same time, although the system has only been utilized to the extent of about one machine search per day. The search products are: (1) a preliminary list of the search terms and the numbers of the selected documents, printed in such a manner that the actual printout can be used as a document request form (see Fig. 7-18); and (2) an abstract of each selected document, printed one per

```
                    D F BERG
                    K102, X2215
                    06/21/62

                    GROUP  0000BA

 CLASSIFICATION

    OF REPORT-

 ACCESS NUMBER.  024401
        SOURCE.  RAND CORP RM-1710
          DATE.  03/22/56
       AUTHORS.  CARTANIO, T F / DREYFUS, S E

                    TITLE AND ABSTRACT

 APPLICATION OF DYNAMIC PROGRAMMING TO THE AIRPLANE MINIMUM TIME-TO-CLI
  PROBLEM.  DYNAMIC PROGRAMMING OFFERS METHOD OF SOLUTION FOR CLIMB
 PROBLEM WHICH IS FAST, READILY ADAPTABLE TO ROUTINE ENGINEERING CALCUL
 ATION AND ALLOWS INCLUSION OF EFFECTS OF VARIATIONS IN AIRPLANE WEIGHT
 AND DRAG ALONG FLIGHT PATH.

 PREPARED BY  FPD TECHNICAL INFORMATION CENTER
```

FIG. 7-19 Sample computer search printout showing abstract of selected document.

page to facilitate screening and filing (see Fig. 7-19). The abstracts are stored on magnetic tape in numerical order, with 10,000 abstracts per tape. Each of the printout sheets is pre-addressed to simplify mailing within the company. The search results are checked by an information specialist, who then groups the abstracts into some order according to their probable relevance. The requester can then review the abstracts, and use a printed list of document numbers as an order form to request the original documents from the library. The present cost to perform retrospective searching with this system ranges from $22 to $115 per customer, depending upon the number of questions that are to be run concurrently. To estimate the document input processing costs, the following figures were presented: indexing and abstracting time averaged 16.5 minutes per document; clerical processing time averaged 7.5 minutes per document; the keypunching was done with one card required for each key word and each accession number used for searching or posting, and 5 to 6 cards per document were required for the abstracts and bibliographic data. After several years of experience with this system, its operators have become increasingly convinced that the major application for their system is for the selective dissemination of information, with retrospective searching as a by-product. They are currently providing a current-awareness service within their own organization for a charge of $25 per month for semi-monthly searches of incoming material. The present program can now search against the complex interest profiles of 20 or more users for less than $80 per machine run. Other by-products of this computer searching system are a manual version of the computer index that is printed and updated by the computer, a program to assist in the preparation and printing of the thesaurus for this file, a weekly accession list printed by the computer in either a subject category or KWIC format, a machine-printed alphabetical listing of the indexing vocabulary that shows the frequency of indexing use of each term, and a descriptor correlation report that shows for any given term all other terms that have been used with it (and the number of times that this joint usage has occurred). An alphabetic listing of the key word dictionary is made periodically that shows the density of posting for each word. These tools permit the user to plan his search strategy to obtain a more selective search product, and provides useful information for the improvement and maintenance of the indexing vocabulary.[62]

Another IBM 704 literature-searching system is in operation at the Allison Division of General Motors Corporation. This system uses a coordinate indexing system and furnishes complete bibliographic data and a brief abstract as a search product. For searching, the key words used to describe the inquiry are punched into cards along with an identification card that contains the requester's name, department number, and date of request. Other search constraints such as allowable ranges of dates, or specific authors, may also be punched and entered as part of the search prescription. The key words are punched into the cards in an equation form that will produce a combination of specified logical operations.[63]

Relatively small computers, such as the Bendix G-15, have also been used for literature searching. One such system at the duPont Company in Wilmington, Delaware, has used such a machine since early 1959 for searching an index to some 14,000 technical reports. An index file is stored on magnetic tape, with all of the subject codes (up to 100) for a single report stored together with the report identification. Up to 16 questions may be asked during one pass through the tape, and each question may involve up to 16 subject codes. The search product is a list of selected document numbers. The search requires approximately 3 hours of machine time for each group of 16 questions. Preliminary data for 501 questions suggest a cost of about 10 dollars to answer the average question, about one-fourth of this going for machine rental, and the rest for the time of the information specialist.[64]

Some programing research studies by Horty and others at the University of Pittsburgh have led to the development of a comprehensive programing system for the analysis and processing of legal text material.[65] Programs have been developed for file construction, searching, concordance construction, KWIC indexing, and statistical studies using an IBM 7070 and 1401 computer. The searching program currently operates on a file of the complete text of the Pennsylvania statutes. The New Jersey court rules, rules of evidence and constitution, were placed on tape under contract from the New Jersey Supreme Court. Searches run on these materials will

[62] Dennis, B. K., "High-Speed Literature Searching on an IBM 704," in *General Information Manual: Information Retrieval Systems Conference, September 21–23, 1959, Poughkeepsie,* Brochure E20-8040 of the IBM Corp., Data Processing Div., White Plains, New York.

Barton, A. R., V. L. Schatz, and L. N. Caplan, "Information Retrieval on a High-Speed Computer," *Proceedings of the 1959 Western Joint Computer Conference,* p. 77 (Institute of Radio Engineers, New York).

Dennis, B. K., "Rapid Retrieval of Information," *Computers and Automation,* Vol. 7, No. 10, pp. 8–9 (October 1958).

Dennis, B. K., J. J. Brady, and J. A. Dovel, Jr., "Five Operational Years of Inverted Index Manipulation and Abstract Retrieval by an Electronic Computer," *Journal of Chemical Documentation,* Vol. 2, No. 4, pp. 234–242 (October 1962).

Dennis, B. K., "Financing a Technical Information Center," pp. 61–75 in *Information Retrieval Management,* L. H. Hattery and E. McCormick, editors (American Data Processing, Inc., Detroit, 1962).

Dennis, B. K., "General Electric's Automatic Retrieval System," paper presented at the Special Libraries Association meeting, Columbus, Ohio, April 1961.

[63] Milligan, J. L., "Information Retrieval for a Technical Library," paper presented at the 15th National Conference of the Association for Computing Machinery, Milwaukee, Wisconsin (August 1960).

[64] Grandine, J. D., E. M. Starr, and R. E. Putscher, "Report Index Searching on the Bendix G-15D Computer," paper presented at the September 1959 American Chemical Society Meeting at Atlantic City.

[65] Kehl, W. B., et al., article cited earlier in this chapter, fn. 7.

be used to aid the court in revising the court rules as well as for finding the various other rules. The search prescription includes single words, groups of synonyms, or collections of either of these, with specified inter-relationships. Thus the words *house, houses, building, buildings, edifice,* and *edifices* might be included in the prescription as a synonym group. The words and synonyms are taken from a vocabulary list, prepared as a part of the file preparation process, which shows the words that have been encountered to date in the file material as well as the frequency of occurrence of each word. The search product is a printout (see Fig. 7-20) of the terms used, the relevant statute citations, and the actual text of the selected citations. The problem posed in Fig. 7-20 was to find those rules which related to the dismissal of appeals for want of prosecution. The search produced 11 documents, of which the first two are shown. The actual text material (6.2 million words) was initially punched onto 635,000 cards, and later copies onto 4 reels of computer magnetic tape. Other preliminary experiments have been made in the searching of natural language text material.[66]

Since January 1960, the Center for Communication and Documentation Research at Western Reserve University, under the joint sponsorship of the National Science Foundation and the American Society for Metals, has operated an experimental machine literature-searching service utilizing, since mid-1961, a GE-225 computer.[67] The file of this system consists of a collection of over 85,000 coded "telegraphic" abstracts of technical documents primarily in the field of metallurgy, and is growing at a rate of several thousand abstracts per month. For a fee, this Metals Documentation Service provides both retrospective searching of the entire file and current-awareness searching of the incoming file material. The search program examines a magnetic tape file (about 17 reels of tape for the present file) for as many as 100 simultaneous searches, each of which may vary in indexing depth up to a limit of 28 levels of logic. The runs to date have seldom searched for more than 25 questions simultaneously for selective dissemination programs and 5 questions simultaneously for retrospective searching. A search of the entire file would take approximately 8 hours of computer time. The user directs his question to a search specialist at the Center skilled in the subject field and the file coding system, who translates it into the appropriate codes and logical symbolization. Each file item is indexed by an average of about 40 concepts or access points. The search program for each user is given a trial run on a single representative reel of tape. The search product is a list of relevant abstract numbers. This list is then used to manually pull copies of the abstracts from a manual file system to be photocopied and delivered to the inquirer. The abstracts are those published in the American Society of Metals *Review of Metal Literature*, or are written in that form if not published. The charge for this service is 50 dollars per month for current-awareness searching, and 200 dollars per year covered for retrospective searches.

A coordinate indexing search program for the IBM 704 has been written and used by the U.S. Naval Ordnance Laboratory in White Oak, Maryland, since late 1960. This particular program uses superimposed coding of the descriptors to reduce the search time and the amount of tape storage required. A search is initiated by reading request cards and the program into the computer. The request cards include such information as the date, the name of the requester, any limitations on dates of interest of the file material, security classification, and no more than 7 search descriptors. The documents are located by serially scanning the magnetic tape files, examining each file item to see if it satisfies the search request. In this system, an average of 11 computer words of 36 bits each is required to describe each document. This description includes the accession number, the originating agency, the date, the security classification, and the

[66] Swanson, D. R., "Searching Natural Language Text by Computer," *Science*, Vol. 132, No. 3434, pp. 1099–1104 (October 21, 1960).

Yngve, V. H., "The Feasibility of Machine Searching of English Texts," *Proceedings of the International Conference on Scientific Information*, Vol. 2, pp. 975–995 (National Academy of Sciences, Washington, D.C., 1959).

[67] Hyslop, M. R., "Machine Literature Searching—From Experiment to Experience," *American Documentation*, Vol. 12, No. 1, pp. 49–52 (January 1961).

Hyslop, M. R., "Metallurgical Documentation from Research to Practice," *Journal of the Franklin Institute*, Vol. 270, No. 1, pp. 27–33 (July 1960).

Kent, A., "Exploitation of Recorded Information: I, Development of an Operational Machine Searching Service for the Literature of Metallurgy and Allied Subjects," *American Documentation*, Vol. 11, No. 2, pp. 173–188 (April 1960).

Perry, J. W., and A. Kent, *Tools for Machine Literature Searching* (Interscience Publishers, New York, 1958).

```
*  HEALTH LAW CENTER
*  UNIVERSITY OF PITTSBURGH
*  INQUIRY NO. 20603-1
*  QUESTION FRAMED BY SCHAFER
*  QUESTION.*  RULES RELATING TO THE DISMISSAL OF
*              APPEALS FOR WANT OF PROSECUTION
*  1/22/63
*
  GROUP 1     OR                         PROSECUTION
              WORD       -2              WANT
              OR                         DEFAULT
*
  GROUP 2     OR                         FAIL
              OR                         FAILED
              OR                         FAILING
              OR                         FAILURE
              OR                         NOT
              SENTENCE                   PROCEED
              SENTENCE                   PROSECUTE
              SENTENCE                   PROSECUTED
              SENTENCE                   PROSECUTING
              OR                                          GROUP 1
              DOCUMENT                   APPEAL
              DOCUMENT                   APPEALED
              DOCUMENT                   APPEALS
*
  GROUP 3     OR                         DISMISS
              OR                         DISMISSAL
              OR                         DISMISSED
              OR                         DISMISSING
              DOCUMENT                                    GROUP 2
              PRINT
* END

N. J. COURT RULE RR 1.*2-10                                                21
RR 1.*2-10.  DEPOSIT FOR COSTS
  IN ALL CIVIL APPEALS THE APPELLANT SHALL WITHIN 30 DAYS AFTER FILING
NOTICE OF APPEAL, DEPOSIT WITH THE CLERK OF THE APPELLATE COURT $100
TO ANSWER THE COSTS OF THE APPEAL, AND IN DEFAULT OF MAKING SUCH DEPOSIT
ON APPLICATION OF THE RESPONDENT THE APPEAL MAY BE DISMISSED BY THE
APPELLATE COURT WITH COSTS., EXCEPT THAT NO DEPOSIT FOR COSTS SHALL BE
REQUIRED WHERE AN APPEAL IS TAKEN BY THE STATE OR ANY POLITICAL
SUBDIVISION THEREOF OR AN OFFICER OR AGENCY OF THE STATE OR ANY POLITICAL
SUBDIVISION THEREOF OR BY DIRECTION OF ANY OF THE PRINCIPAL DEPARTMENTS
OF THE STATE, OR WHERE THE APPELLANT PURSUANT TO RULE 1.*4-8 HAS FILED
A SUPERSEDEAS BOND OR MADE A DEPOSIT IN COURT IN LIEU THEREOF.
NOTE.*  FORMERLY RULE 1.*2-15.  APPLIES TO SUPREME COURT CERTIFICATION.
APPLIES TO ALL APPELLATE COURTS.
HOW APPEALS ARE TAKEN

N. J. COURT RULE RR 1.*4-1                                                 27
RR 1.*4-1.  SUPERVISION OF APPEALS IN ALL COURTS
  THE SUPERVISION AND CONTROL OF THE PROCEEDINGS ON APPEAL SHALL BE IN
THE APPELLATE COURT FROM THE TIME THE APPEAL IS TAKEN, EXCEPT AS OTHERWISE
PROVIDED IN THESE RULES.  SUBJECT TO THE PROVISIONS OF RULE 1.*4-2 IN
CRIMINAL CASES, THE FAILURE OF A PARTY TO PROSECUTE HIS APPEAL SHALL
NOT AFFECT THE VALIDITY OF THE APPEAL, BUT SHALL BE GROUND ONLY FOR
SUCH ACTION AS THE APPELLATE COURT MAY DEEM APPROPRIATE, WHICH MAY
INCLUDE DISMISSAL OF THE APPEAL.  THE APPELLATE COURT MAY AT ANY TIME
ENTERTAIN A MOTION TO DISMISS THE APPEAL, OR FOR DIRECTIONS TO THE COURT
OR COURTS BELOW, OR TO MODIFY OR VACATE ANY ORDER MADE BY THE COURT    ,
OR COURTS BELOW, OR BY ANY JUDGE, IN RELATION TO THE PROSECUTION OF
```

FIG. 7-20 Sample search printout from a legal text searching program.

superimposed code pattern. Since a standard 2400-foot reel of IBM 704 magnetic tape can store approximately 900,000 of these computer words, it is possible to store the index to approximately 80,000 documents on a single reel of tape. A coarse screening is first performed by checking to see if the quiz pattern is included in the superimposed index pattern of the document. This is a relatively rapid test, and rejects most of the documents. The documents that pass the initial screening *may* be relevant, and so are subjected to a finer screening operation which checks each specific search descriptor against the document's list of specific descriptors. The search product is a list of selected document numbers. A typical request load of 20 queries of a file of 50,000 items would require approximately 1 minute of searching time and about 4 minutes of auxiliary time for such functions as reading the queries and program into the machine. It has been suggested that disregarding the initial programing cost, this would result in a cost of about $1.25 per query for an exhaustive search of this file. The file updating is done at the same time as the searching. The actual program for file updating and searching contained nearly 3000 instructions and required about a year for its preparation and testing.[68]

The Armed Services Technical Information Agency (ASTIA) in Arlington, Virginia, is one of the largest technical information centers in the world. This organization, established to provide a central service within the U.S. Department of Defense for the efficient interchange of scientific and technical information, collects and disseminates copies of technical reports and other documents prepared as a result of Department of Defense research and development contracts. Approximately 250,000 reports are indexed in the file collection of about 650,000 reports, and new reports are being received at a rate of about 30,000 titles per year. Approximately 700,000 reports are furnished to requesters each year, and approximately 5000 requests are received each year for bibliographic searches. A Univac Solid-State 90 computer system was installed in early 1960 to assist in the file-searching operation and to perform other functions. The Uniterm system of coordinate indexing is used to describe the subject matter of the reports with an average of 8 Uniterms per document. A comprehensive technical thesaurus was developed to assist in the control and use of the indexing terms. The development of the first thesaurus preceded the actual punching of the cards for entry into the machine system, which was accomplished in late 1960. With the magnetic tape system it was possible to file 470,400 document numbers under Uniterm headings on a single 2400-foot reel of magnetic tape. Consequently it was possible to put the entire document index to approximately 250,000 ASTIA documents on 4 reels of tape. The actual search program handles as many as 10 searches simultaneously, with a maximum of 4 retrieval term coordinations for any search. The search program reads in punched cards to describe the search prescription, and then searches the master tape file for pertinent document numbers. The selected document numbers are then punched out, one to each card. In current practice, about 8 hours of computer time are required to run 10 average searches employing a total of 120 to 130 different Uniterms.[69]

One other large computer searching system, utilizing a Honeywell H-800 computer, is currently under development at the National Library of Medicine and is expected to become fully operational in 1964. The system will be used to perform retrospective searching of medical literature as well as assist in the preparation of *Index Medicus*, the world's largest index journal.

For manual or machine searching systems, some thought has been given to the idea that an index term should have a range or degree of application to a document, rather than being so used that the index term either applies or does not apply.[70] That

[68] *Pioneering in Machine Literature Searching and Retrieval*, NAVWEPS Report 7388, Proceedings of the Sixth Meeting of the Council of Librarians of the East Coast Navy Laboratories, AD-255 291 (April 1961).

[69] Armed Services Technical Information Agency, Arlington, Virginia, *Automation of ASTIA—1960*, AD-247 000 (December 1960).

Hammond, W., *Evolution of the ASTIA Automated Search and Retrieval System*, AD-252 000 (January 1961).

Barden, W., W. Hammond, and J. Heald, *Automation of ASTIA—A Preliminary Report*, AD-227 000 (December 1959).

Controlling Literature by Automation, AD-243 000 (October 5–7, 1960), Proceedings of the Fourth Annual Military Librarian's Workshop.

[70] Maron, M. E., J. L. Kuhns, and L. C. Ray, *Probabilistic Indexing—A Statistical Technique for Document Identification and Retrieval*, June 1959 Technical Memo No. 3, Data Systems Project Office, Ramo-Wooldridge, Los Angeles, California.

Maron, M. E., and J. L. Kuhns, "On Relevance, Probabilistic Indexing and Information Retrieval," *Journal of the* is, for a given document, give each assigned index term a weight or value between 0 and 1, instead of assuming that each index term is equally descriptive of the document contents. Since index terms can only approximate the contents of the document or the intents of a searcher, weighing each index term will allow the approximation made (by the indexer and the searcher) to be closer to the content of a document. Conventional searching systems usually perform an exact match of the specific search criteria against the indexes of the documents in the collections, to produce a list of file documents which satisfy the search criteria. No indication is usually given as to which of the selected documents are more relevant to the search than others. However, probabilistic indexing as proposed here, would —for a given query—weigh the index terms that are assigned to documents and use these weights to compute a number (called a "relevance number") for each document, thus providing a measure of the probability that the document will satisfy the search request. The result of a search is an ordered list of documents that satisfy the request, ranked according to their probable relevance. This technique can be implemented rather easily by a computer searching system, whereas it would be very cumbersome to do it with manual methods. Additional features of this technique are the ability of the computer to automatically modify the search strategy or to elaborate on the request in the most probable direction in order to increase the probability of selecting relevant documents. Some experiments which compared searching with weighted index terms of values between 0 and 1 to searching with a conventional system that used weights of only 0 and 1 showed that the searches with weighted index terms generally disclosed as many relevant, and fewer irrelevant documents, than the conventional system. Some work has been done along this line by Stiles, in which a coordinate indexing system was used with an IBM 705 to find documents related to a request even though they might not be indexed by the exact terms of the request.[71] The computer presents the user with a list of documents ranked in the order of their relevance to the request. Stiles' method determines the degree of association between pairs of index terms in such a way that a vocabulary of words (an extended list of request terms) closely related to any given term or group of terms can be derived. A comparison is made of the amount of agreement between this extended list of request terms and the terms used to index a document, and the measure of this agreement is used as a measure of the document's relevance to the search request.

In looking back over the described applications and additional references, we find that file-searching operations have been programed for nearly all of the larger general-purpose computers, and with a number of different indexing systems, although most systems used some form of coordinate indexing. Several of the reported installations were experimental, but many systems are operating on a regular production basis. In most cases, the machines used were already available within the organization—few organizations obtained computers solely for information retrieval. With the exception of a few IBM RAMAC systems, the retrieval systems employed computers that used magnetic tape. A recent survey by the National Science Foundation reported 10 different systems in operation for storing and searching files of references with a computer.[72]

Because of the comparatively high cost of programing and operation, computers are most profitable in the following situations: (1) when there is a need for a completely exhaustive search of a large file; (2) when there is a requirement for a relatively rapid response to an inquiry and the computer can be made available on demand; (3) when there is a particularly complex question or file of complex material, such as chemical structures or patents; (4) when a relatively large number of searches must be conducted; and (5) when the field covered is lacking in good bibliographical control in conventional form and does not lend itself readily to such control. With the indexing information now in machine language, the user enjoys the additional advantage of being able to easily obtain statistical data on the use of the file material or classification system, in order to provide corrective action.

Association for Computing Machinery, Vol. 7, No. 3, pp. 216–244 (July 1960).

[71] Stiles, H. E., "Machine Retrieval Using the Association Factor," in *Machine Indexing: Progress and Problems,* pp. 192–206 (American University, Washington, D.C., 1961).

Stiles, H. E., "The Association Factor in Information Retrieval," *Journal of the Association for Computing Machinery,* Vol. 8, No. 2, pp. 271–279 (April 1961). See also the references to Swanson in this chapter, fn. 21, fn. 66.

[72] *Nonconventional Technical Information Systems in Current Use, No. 3,* Report NSF-62-34 of the National Science Foundation, Washington, D.C. (October 1962).

Nearly all of the file-searching systems have been concerned with technical literature such as engineering documents and reports, scientific journal articles, and patents. Searching was performed in "batch" operations, that is, the search requests were allowed to accumulate, and once per day, once per week, or once per other period, the search program was run to answer all the requests at the same time. File maintenance was also done in "batches." This batching operation is necessary to minimize the computer set-up time (and hence the additional charges) required when the computer is changed from one job to another. As much as 5 minutes may be required to get the computer ready for the file search program, regardless of the number of searches to be conducted. Since a batch mode is used, the user may have to wait a day or two for the next run in order to get the answers to his question, even though the computer can search the files at speeds of 10,000 or more items per minute.

Generally, only the indexing information is stored in the computer, and the result of a search is a list of document numbers or titles. In a few systems an abstract is printed out, although the cost in money, memory space, and computer time has usually discouraged the storage of much descriptive information in machine language.

As mentioned previously, a considerable amount of time and money must be spent to develop a program and data base for computer searching. Even if the program is obtained at no cost, and little or no computer time is required for searches, there will still be the continued cost of converting new information into machine language for file updating. This might require the expenditure of something on the order of 50 cents per file item.

With their current configuration and operating cost, it is impractical to use computers for casual browsing of the file. In addition, they are limited to storing information in digital form in a restricted set of characters. There is currently no convenient way to print or store symbolic information such as equations, chemical structures, and special notations. A feature lacking in most of the programing systems is the capability for searching while automatically considering all possible synonyms, to assure greater completeness. Moreover, very little progress has been made in mechanizing an associative or "see also" search.

In many of the earlier literature-searching computer systems reported, the original manual system could probably have performed as well or better, and at less cost. The computer systems may have contributed speed, accuracy, additional printout, and growth potential, but it is highly probable that the decision to try a computer system was influenced in large measure by curiosity and its novelty, and not by its economics. In few instances is it clear that the computer system did anything more than mechanize a manual or punched card system. The computer file parameters, such as the number of descriptors per document, and the number of descriptors in the dictionary, were generally the same as in a manual system. The details of the programing logic and the organization of data in the machine and on the tapes differed from those in the manual systems, but these details were usually dictated by the characteristics of the machine being used. In most cases, the computer system operated in addition to a manual system, instead of replacing it, which is probably the way it should be operated. The inaccessibility of the computer for quick inquiries and the need for a browsing capability demands that an additional manual file be maintained. Under these circumstances, a computer system can be expected to add to the library's operating costs, instead of reducing them—although there may be an improvement in service. Costs ranging from 1 to 100 dollars per question for machine searching have been mentioned, but this figure is subject to great variation, depending upon the particular computer, the size of the file, and the basis for figuring the costs.

In addition to literature searching, the searching of chemical compounds and structures was the subject of much of the early work in computer searching. Whereas the mechanized literature-searching systems resembled contemporary manual systems, new and original work was done to mechanize chemical searching. Special coding techniques made it possible to describe the composition and topography of the compounds in language adapted to the computers. Several codes were evolved for this purpose, and programs and sophisticated input-output procedures were developed for several different computers. In chemical-structure searching it is decidedly more convenient to use a computer instead of a manual system. The search operation is so complex, the data base so large, and the available manual reference tools so inadequate that computer searching is the only really satisfactory method.

Automatic Dissemination and Routing of Information

Many information officers or librarians in many organizations have provided a manual dissemination and sorting system for many years, keeping track of their clients' interests, and sending them material that they feel is pertinent. In those cases where a requester's interests can be represented in the notation used to describe documents or other items of information, it is possible to mechanically match the interests to the documents. This is no different from stating a person's interests as a search question, and then continuously searching all the new information as it is entered into the file system. One of the earliest experiments along these lines was an effort by Luhn to selectively and automatically disseminate information about new documents and library acquisitions to research workers who had a potential interest in these materials.[73] The interests of the workers were characterized by an "interest profile," consisting of a list of weighted key words or descriptors (normally from 2 to 26 key words). The descriptors of incoming documents were compared with those in groups of interest profiles to find out who might be interested. A notification card with an abstract of the document (see Fig. 7-21) was then sent to potentially interested persons. The portion of the card that contained an abstract of the document could be kept by the recipient for a manual card file and the remaining portion of the card could be sent in as request for a copy of the document.

No dissemination system can operate properly unless there is provision for a feedback of information from the user to provide guidance to the system. Luhn built feedback into his system by noting whether or not the user returned the stub from the notification card. This stub was used as a basis for the following actions: (1) if the document was of interest (i.e., if the stub was returned), the entire list of key words for that document was added to the interest profile of that individual, thus confirming or reinforcing old interests and introducing possible new interests; (2) if the document was not of interest (i.e., if the stub was not returned), the entire list of key words for that document was subtracted from the individual's interest profile.

The pilot system, SDI-1, was tested with an IBM 650 in 1959. SDI-2 has recently reached fully operational status at the IBM Advanced Systems Development Division in New York, to serve approximately 500 users. The computer programs for the SDI-1 system are available for the IBM 650 on any data processing system that can use a program written in FORTRAN II language. The SDI-2 system can be run on an IBM Tape 650, 704, 709, 7090, and 7094. The SDI-3 system can be run on the IBM Tape 1401.[74] During the first year of operation at IBM, approximately 21,000 documents were processed. Close to 13,000 notices were sent out in a recent 6-month period, with a positive acceptance of 8000 cases, and a resultant 4500 requests for the original documents. Preliminary studies suggest that a Selective Dissemination of Information (SDI) system for about 1000 users with an input of about 500 documents per month would cost from 25 to 50 dollars per user per year.

Copy Editing

The writing functions will see increasing automation of their abstracting and indexing operations. This will include the compilation of bibliographies, the composition of title and abstract journals, and the production of other lists and directories. Automation techniques will be used to prepare indexes for manuscripts prior to publication. And in addition, concordances—which have been used to date primarily by academicians—will start being used in abbreviated form for such purposes as guides to textbooks, instruction and procedures manuals, tax guides, and legal literature. Some work has even

[73] Luhn, H. P., "Business Intelligence System," *IBM Journal,* Vol. 2, No. 4, pp. 314–319 (October 1958).

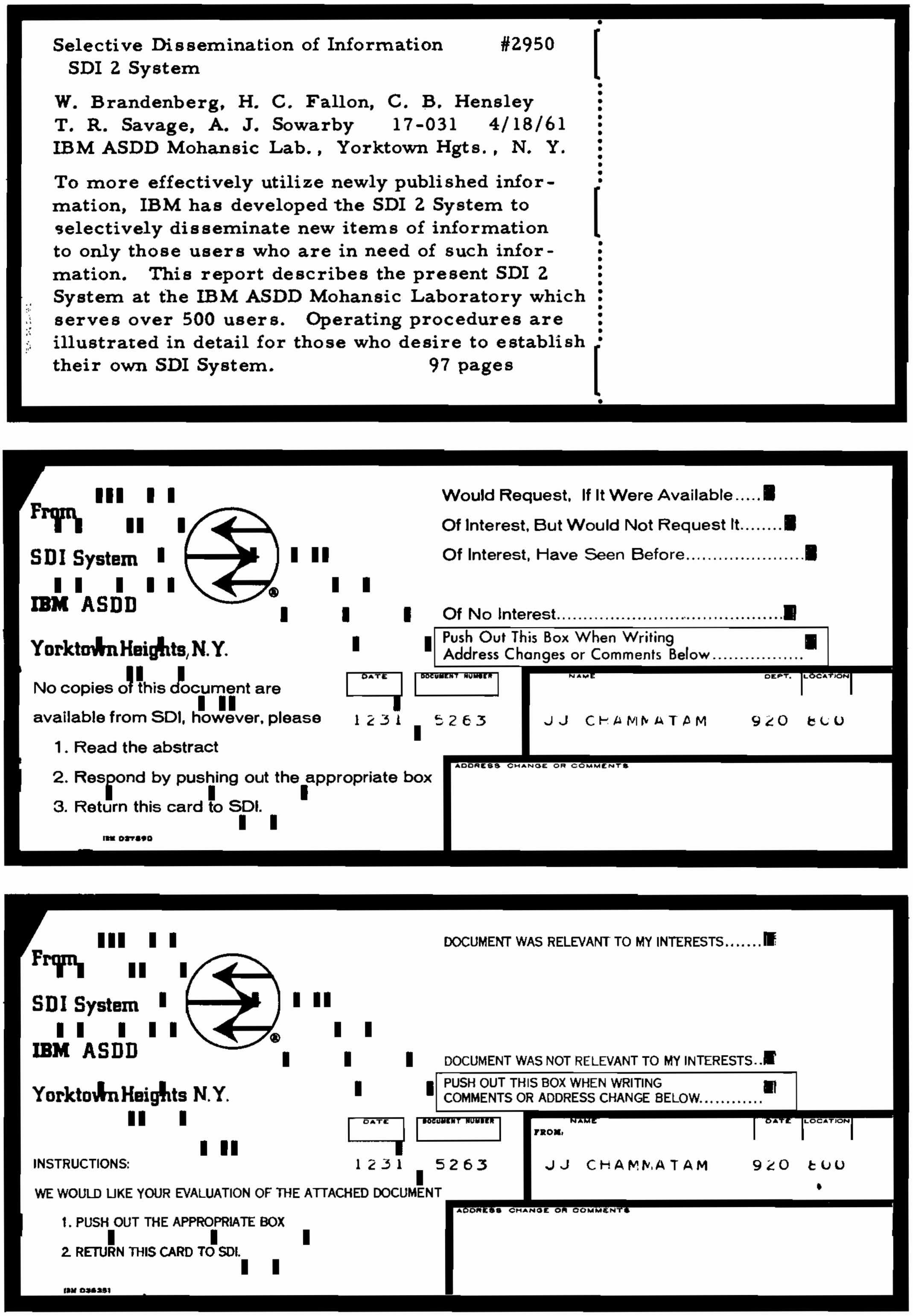
Luhn, H. P., *Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment,* an un-numbered report of the IBM Advanced Systems Development Div., Yorktown Heights, New York (November 1959).

Luhn, H. P., "Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment," *American Documentation,* Vol. 12, No. 2, pp. 131–138 (April 1961).

General Information Manual: Selective Dissemination of Information, Brochure E20-8092 of IBM Corp., Data Processing Div., White Plains, New York (1962).

[74] Brandenberg, W., et al., *Selective Dissemination of Information SDI-2 System,* Report 17-031 of the IBM Corporation, Advanced Systems Development Div., Yorktown Heights, New York (April 1961).

Hensley, C. B., et al., "Selective Dissemination of Information—A New Approach to Effective Communication," *Institute of Radio Engineers Transactions on Engineering Management,* Vol. EM-9, No. 2, pp. 55–65 (June 1962).



Selective Dissemination of Information #2950
SDI 2 System

W. Brandenberg, H. C. Fallon, C. B. Hensley
T. R. Savage, A. J. Sowarby 17-031 4/18/61
IBM ASDD Mohansic Lab., Yorktown Hgts., N. Y.

To more effectively utilize newly published information, IBM has developed the SDI 2 System to selectively disseminate new items of information to only those users who are in need of such information. This report describes the present SDI 2 System at the IBM ASDD Mohansic Laboratory which serves over 500 users. Operating procedures are illustrated in detail for those who desire to establish their own SDI System. 97 pages

From
SDI System
IBM ASDD
Yorktown Heights, N.Y.

Would Request, If It Were Available.....
Of Interest, But Would Not Request It........
Of Interest, Have Seen Before....................
Of No Interest..
Push Out This Box When Writing Address Changes or Comments Below................

No copies of this document are available from SDI, however, please
1. Read the abstract
2. Respond by pushing out the appropriate box
3. Return this card to SDI.

DATE	DOCUMENT NUMBER	NAME	DEPT.	LOCATION
1231	5263	JJ CHAMMATAM	920	600

ADDRESS CHANGE OR COMMENTS

IBM 027890

From
SDI System
IBM ASDD
Yorktown Heights N.Y.

DOCUMENT WAS RELEVANT TO MY INTERESTS.......
DOCUMENT WAS NOT RELEVANT TO MY INTERESTS..
PUSH OUT THIS BOX WHEN WRITING COMMENTS OR ADDRESS CHANGE BELOW...........

INSTRUCTIONS:
WE WOULD LIKE YOUR EVALUATION OF THE ATTACHED DOCUMENT
1. PUSH OUT THE APPROPRIATE BOX
2. RETURN THIS CARD TO SDI.

DATE	DOCUMENT NUMBER	FROM: NAME	DATE	LOCATION
1231	5263	JJ CHAMMATAM	920	600

ADDRESS CHANGE OR COMMENTS

IBM 036351

FIG. 7-21 Representative notification-abstract cards.

been reported recently on computer programs to systematically generate lists of pronounceable three- and four-letter words.[75]

There is a very real possibility of initiating some mechanization of the editing function. It is possible to perform many copy-editing functions by a table look-up operation (e.g., standard forms for abbreviations and journal citations, correct spelling, approved terminology). Consequently, some of these functions could be mechanized. It should also be possible to do some editing (for form, not content) with data processing equipment in much the same manner as that employed in machine translation, using rules of grammer and word dictionaries. Some computer programs have been written to perform some editing and composing of text material, along with a subsequent conversion into a code suitable for a photocomposing machine.[76]

Several research projects have investigated the possibility of performing some of the copy-editing functions with computers, by doing such things as automatically extrapolating and filling in missing words in text material, correcting misspelled words, and automatically inserting spaces where words had been run together.[77] In one instance, 4200 words of text material that had been squeezed together were expanded automatically back into text form with a correction rate of 99.7 percent.[78] Andifyoustoptothinkaboutitthisisaprettydarngoodrate.

[75] Leiner, A. L., and W. W. Youden, "A System for Generating 'Pronounceable' Names Using a Computer," *Journal of the Association for Computing Machinery,* Vol. 8, No. 1, pp. 97–103 (January 1961).

[76] Barnett, M. P., et al., "Computer Preparation of Photocomposing Control Tapes. Part 1. Preparation of Flexowriter Source Material," *American Documentation,* Vol. 13, No. 1, pp. 58–65 (January 1962).

Yasaki, E., "The Computer and Newsprint," *Datamation,* Vol. 9, No. 3, pp. 27–31 (March 1963).

[77] Blair, C. R., "A Program for Correcting Spelling Errors," *Information and Control,* Vol. 3, No. 1, pp. 60–67 (March 1960).

Miller, G. A., and E. A. Friedman, "The Reconstruction of Mutilated English Texts," *Information and Control,* Vol. 1, No. 1, pp. 38–55 (September 1957).

Chapanis, A., "The Reconstruction of Abbreviated Printed Messages," *Journal of Experimental Psychology,* Vol. 48, No. 6, pp. 496–510 (December 1954).

Evans, M. W., et al., *Machine Correction of Garbled English Text,* Report 54G-0022 of the Massachusetts Institute of Technology, Lincoln Laboratory (June 1960), AD-237 700.

Glantz, H. T., "On the Recognition of Information with a Digital Computer," *Journal of the Association for Computing Machinery,* Vol. 4, No. 2, pp. 178–188 (April 1957).

Drafting and Related Functions

A tremendous number of drafting and engineering man-hours are spent annually throughout the country in the manual preparation of engineering drawings, blueprints, schematic diagrams, flow charts, and similar documents. A few innovations have been used to augment but not replace the manual process—special templates, pre-printed diagrams of common components, simplified drawing methods, and so on. Computer programs have been used in a few isolated instances to perform some of these drafting tasks.

For organizations that write a large number of complex computer programs, considerable drafting work must be done to draw the flow charts that describe the operations of these programs. One flow chart may run to scores of pages for a single program, and can be extremely complicated. Several successful computer programs have been written to produce flow charts or block diagrams automatically from a given list of program instructions.[79] The programs even go so far as to label each box to show in summary fashion what variables were involved, what computations were performed, and what input-output equipment was used.

In industries where numerically controlled machine tools are used, the description of a particular part design may take the form of some equations and geometrical statements. Computer programs are available to take this descriptive information and generate the sequence of commands for the machine tool. In addition, programs are also available to take these geometrical or mathematical statements and generate an actual image or engineering drawing on a cathode ray tube display

[78] Evans, M. W., et al., "Degarble," a portion (p. 35) of the report, *Information Processing. Quarterly Progress Report of Division 5,* Massachusetts Institute of Technology, Lincoln Laboratory (June 15, 1960).

[79] Haibt, L. M., "A Program to Draw Multilevel Flow Charts," *Proceedings of the 1959 Western Joint Computer Conference,* pp. 131–137 (The Institute of Radio Engineers, New York).

Scott, A. E., "Automatic Preparation of Flow Chart Listings," *Journal of the Association for Computing Machinery,* Vol. 5, No. 1, pp. 57–66 (January 1958).

device for subsequent photographing and printing.[80]

An extensive amount of work has been done in the computer industry to completely automate the major portions of the engineering documentation functions that accompany the design and development of a digital computer.[81] The decision to automate seems to stem from the fact that a large-scale computer is simply too complex to document and control with manual methods, and that manual methods are too slow to be tolerated in an industry where the products are so subject to technical obsolescence.

Design automation procedures in the computer industry incorporate the following types of semiautomatic functions that are concerned with project documentation and associated clerical tasks:

Assign all wiring points, component locations, and other fabrication information to the schematic drawings.

Prepare lists of parts locations and wiring information in such a form that an assembler can use this list instead of a drawing.

Prepare system block diagrams from the logical equations.

Route the wiring and prepare a wiring chart and cable listing in such a way as to minimize lead length, while at the same time considering loading and noise effects.

Accommodate engineering changes to make updated lists and diagrams.

Generate bills of materials for electrical components from the logic equations.

Print logic block diagrams from the design information stores on magnetic tape.

Prepare location charts for the placement of circuit cards.

Check some types of design errors.

Use the design information to generate cards or tapes for direct input into automated fabrication equipment.

All of these tasks are performed semi-automatically by the computer programs. However, this design automation process is costly; some complete systems require as many as 120,000 program instructions. Nevertheless, many of the technical drafting, writing, and editing chores involved have been automated to the apparent satisfaction of the manufacturers.

As mentioned earlier in this chapter, the automatic generation, display, and printing of diagrams of chemical structures has also been accomplished with data processing equipment. Computer programs have also been written to provide the necessary spacing and justification of the symbols in an equation or chemical structure diagram so that a paper tape can be punched for a high-quality-type composing machine.[82]

REPRESENTATIVE COSTS

Computer equipment, whether bought or rented, is so expensive that only a few organizations (e.g., ASTIA and the National Library of Medicine) have been able to justify installing it for the primary purpose of handling storage and retrieval problems. The majority of firms using computers for information retrieval have used machines at service bureaus, or shared company machines that were purchased primarily for some other purpose. Computer time may be purchased on medium and large computers at rates of around 50 to 550 dollars per hour. The purchase price or rent of a particular system depends upon the system configuration, e.g., the memory size, the number of tape transports, and the number of special features. Figure 7-22 illustrates the range of equipment costs for many of the available large and medium-size computer systems.

For any data processing system, the cost of the computer is only a fraction of the total cost to the user. The cost of installing and running a computer

[80] Anon., "Automatic Drafting Moves Closer," *Control Engineering*, Vol. 8, No. 2, pp. 31–32 (February 1961).

[81] Kloomok, M., et al., "The Recording, Checking, and Printing of Logic Diagrams," *Proceedings of the 1958 Eastern Joint Computer Conference*, pp. 108–118 (Institute of Radio Engineers, New York, 1958).

Rocket, F. A., "A Systematic Method for Computer Simplification of Logic Diagrams," *1961 IRE National Convention Record* (Institute of Radio Engineers, New York, 1961).

Cray, S. R., and R. N. Kisch, "A Progress Report on Computer Applications in Computer Design," *Proceedings of the 1956 Western Joint Computer Conference*, pp. 82–85 (Institute of Radio Engineers, New York, 1956).

Milbrain, J. P., and A. V. Banes, "Automated Computer Design," paper presented at the 14th National Association for Computing Machinery, Boston, Massachusetts (September 1959).

Connolly, T. A., "Automated System and Logical Design Techniques for the RW-33 Computer System," *1960 IRE Convention Record*, Part 2 (Institute of Radio Engineers, New York, 1960).

[82] Feldman, A., and H. L. Manceaux, "An Improvement in the Printing of Chemical Structures, Which Results in Their Complete Computer Codes," paper presented at the annual convention of the American Documentation Institute, December 14, 1962, Hollywood-by-the-Sea, Florida.

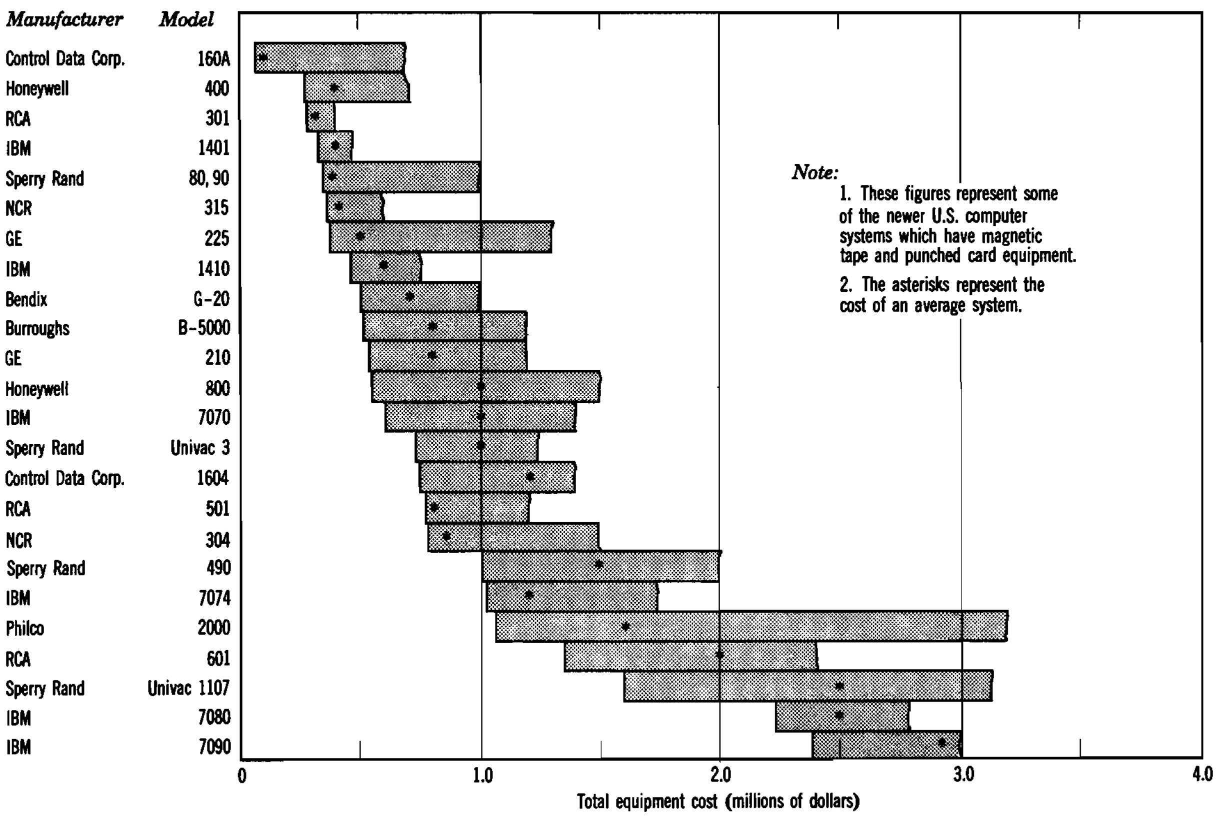


FIG. 7-22 Approximate price ranges of representative computer equipment.

seems to range from two to five times the cost of the equipment, regardless of the computer model, the type of use, and whether the equipment is purchased or leased. Relatively large fixed costs (site preparation and installation, training, initial programing file conversion, etc.) and recurring costs (operating costs, programing, and maintenance, facilities) face any computer installation. The writing and de-bugging of computer programs are both expensive and time-consuming. These large programing costs are part of the motivation for documenting and exchanging programs with other users. A few information retrieval programs are available from various computer manufacturers and cooperative user groups. Some of these were listed earlier, in Table 7-1. As mentioned previously in this chapter, a survey of available programs specifically for chemical information searching has recently been made by the National Bureau of Standards.[83]

[83] Marden, E. C., and H. R. Koller, fn. 59.

Additional References

American University, *Machine Indexing: Progress and Problems,* collection of papers presented at the Third Institute on Information Storage and Retrieval at American University, Washington, D.C., February 1961.

Anon., *Application Brief: Information Retrieval at the U.S. Army Logistics Management Center with the IBM* RAMAC *305,"* Brochure K50-0006 of the IBM Corp., Federal Systems Div., Rockville, Maryland (1961).

Black, D., "Large Academic Libraries and Data Processing," *Data Processing Digest,* June 1962, pp. 21–23; July 1962, pp. 20–22.

Bourne, C. P., "The Beginnings of Automation of Technical Drafting, Writing, and Editing Functions," *Proceedings of the Eighth Annual Convention of the Society of Technical Writers and Publishers,* San Francisco (April 1961).

Costello, J. C., Jr., "Computer Requirements for Inverted Coordinate Indexes," *American Documentation,* Vol. 13, No. 4, pp. 414–419 (October 1962).

Faden, B. R., "Information Retrieval on Automatic Data Processing Equipment," *Special Libraries,* April 1959, pp. 162–165.

Freed, R. N., "A Lawyer's Guide Through the Computer

Maze," *The Practical Lawyer*, Vol. 6, No. 7, pp. 15–45 (November 1960).

Frome, J., "Semi-Automatic Indexing and Encoding," *Journal of the Franklin Institute*, Vol. 270, No. 1, pp. 3–26 (July 1960). (Adapted from U.S. Patent Office Research and Development Report No. 17.)

Gilman, L., and C. Jennings, "An Information Retrieval System Tailored to the Needs of an Electronic Engineering Organization," *National Electronics Conference Proceedings*, Vol. 17, pp. 445–452 (1961).

Gotlieb, C. C., and J. N. P. Hume, *High-Speed Data Processing* (McGraw-Hill Book Co., New York, 1958).

Gurk, H. M., and J. Minker, "The Design and Simulation of an Information Processing System," *Journal of the Association for Computing Machinery*, Vol. 8, No. 2, pp. 260–270 (April 1961).

Kokie, J. E., "An Experimental Numerical Information Retrieval System," paper presented at the *15th National Conference of the Association for Computing Machinery*, Milwaukee, Wisconsin (August 1960).

Ledley, R. S., *Programing and Utilizing Digital Computers* (McGraw-Hill Book Co., New York, 1962).

McCracken, D. D., *A Guide to* ALGOL *Programing* (John Wiley and Sons, New York, 1962).

McCracken, D. D., *A Guide to* FORTRAN *Programing* (John Wiley and Sons, New York, 1957).

McCracken, D. D., *Digital Computer Programing* (John Wiley and Sons, New York, 1957).

Miller, L., et al., "A Multi-Level File Structure for Information Processing," *Proceedings of the 1960 Western Joint Computer Conference*, pp. 53–59 (Institute of Radio Engineers, New York).

Minsky, M., "Descriptive Languages and Problem Solving," *Proceedings of the 1961 Western Joint Computer Conference*, pp. 215–218.

Moore, R. T., "A Screening Method for Large Information Retrieval Systems," *Proceedings of the 1961 Western Joint Computer Conference*, pp. 259–274.

Moyer, S. R., "Automatic Search of Library Documents," *Computers and Automation*, Vol. 6, No. 5, pp. 24–29 (May 1957).

Nanus, B., "The Use of Electronic Computers for Information Retrieval," *Bulletin of the Medical Library Association*, Vol. 48, No. 3, pp. 278–291 (July 1960).

Needham, R. M., "A Method for Using Computers in Information Classification," *Information Processing 1962*, pp. 284–287, C. M. Popplewell, Ed. (North Holland Publishing Co., Amsterdam, 1963).

O'Connor, J., "Some Suggested Mechanized Indexing Investigations Which Require No Machines," *American Documentation*, Vol. 12, No. 3, pp. 198–203 (July 1961).

Opler, A., and N. Baird, "Relative Merits of General and Special Purpose Computers for Information Retrieval," *Proceedings of the 1959 Western Joint Computer Conference*, pp. 54–56 (Institute of Radio Engineers, New York, 1959).

Opler, A., and N. Baird, "Experience in Developing Information Retrieval Systems on Large Electronic Computers," *Proceedings of the International Conference on Scientific Information*, Vol. 1, pp. 699–710 (National Academy of Sciences, Washington, D.C., 1959).

Pantyukhina, M. Ye., "Machine Input and Output of Information on the Structure of Chemical Compounds," *Foreign Developments in Machine Translation and Information Processing*, No. 31, JPRS: 8479, distributed by Office of Technical Services, U.S. Department of Commerce, Washington, D.C. (June 22, 1961).

Pappe, I. Ya., "The Problem of the Automatic Compilation of Organic Compound Indexes for Chemical Abstract Journals," *Foreign Developments in Machine Translation and Information Processing*, No. 31, JPRS: 8479, distributed by Office of Technical Services, U.S. Department of Commerce, Washington, D.C. (June 22, 1961).

Schultz, C. K., and C. A. Shepherd, "A Computer Analysis of the Merck Sharp and Dohme Research Laboratories Indexing System," *American Documentation*, Vol. 12, No. 2, pp. 83–92 (April 1961).

Schultz, C. K., and P. A. Schwartz, "A Generalized Computer Method for Index Production," *American Documentation*, Vol. 13, No. 4, pp. 420–432 (October 1962).

Spiegelthal, E. S., "Computing Educated Guesses," *Proceedings of the 1959 Western Joint Computer Conference*, p. 70 (Institute of Radio Engineers, New York, 1959).

Wright, M. A., "Mechanizing a Large Index," *The Computer Journal* (British), Vol. 3, No. 2, pp. 76–83 (July 1960).

Wright, M. A., "Matching Inquiries to an Index," *The Computer Journal* (British), Vol. 4, No. 1, pp. 38–41 (April 1961).

Wrubel, M. H., *A Primer of Programing for Digital Computers* (McGraw-Hill Book Co., New York, 1959).

8

Other paper tape and magnetic media equipment

PAPER TAPE EQUIPMENT

Punched paper tape is coming into more common use for temporary information storage, for point-of-action data acquisition recording, and as an input medium for information processing equipment. Its popularity is due in large measure to the fact that the punching and reading equipment is relatively simple and inexpensive and provides a permanent storage medium. The tape is most commonly used in conjunction with special electric typewriters, either to actuate the typewriter or to record the operator's typing. Most of the equipment and applications that are relevant to this book make use of the tape in this manner. That is, data are punched in paper tape at the point of origin as a by-product of writing a document, and the tape is subsequently used at some later date to furnish the original data in a machine language. The specific techniques for coding the tape were described in Chapt. 4.

One of the limitations to the use of paper tape with electric typewriters is the difficulty of correcting errors or making changes in the tapes. For example, a typing error on a conventional typewriter can, in most cases, be corrected eimply by striking over the character or backspacing to type a character in a spot that was left as a space or blank. These methods cannot be used with paper tape equipment. Usually the erroneous punches are noted as soon as they happen and are immediately over-punched by the operator with an "ignore" or "delete" pattern, which causes that particular pattern to be ignored by the equipment during subsequent readings. If an error is later discovered to be in the middle of a typed sequence, a new tape must usually be prepared. The common procedure in this case is to duplicate (automatically) a new tape up to the point of the error, type in the correct character, and then copy the data that followed the error. The copy or duplicate operations are performed automatically on most of the paper tape typewriter units.

The paper tape units usually type a hard copy at the same time that the tape is being punched. This provides a visual check on the information being punched. A check for accuracy can be made by running the tape back through the machine to generate another hard copy printout for comparison with the original hard copy. The typewriter types a character or performs a function (e.g., carriage return) to correspond to each of the characters punched on the tape.

For small amounts of punching, information can usually be entered in paper tape form quicker than in punched card form. However, for longer jobs (e.g., several hours or days) that require considerable accuracy, the paper tape operation tends to be slower than the transcription into punched cards, primarily because more time is required to correct mistakes on paper tape than on punched cards.

The Equipment

Basic Typewriter Equipment. The Friden Flexowriter (Fig. 8-1) and the Remington Synchro-Tape

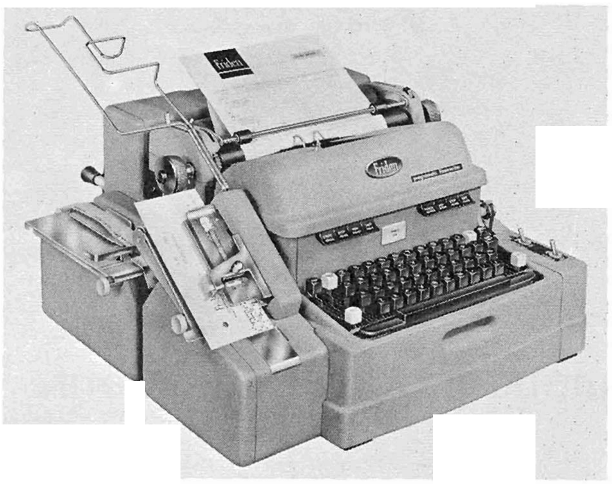


FIG. 8-1 Friden Programatic Flexowriter.

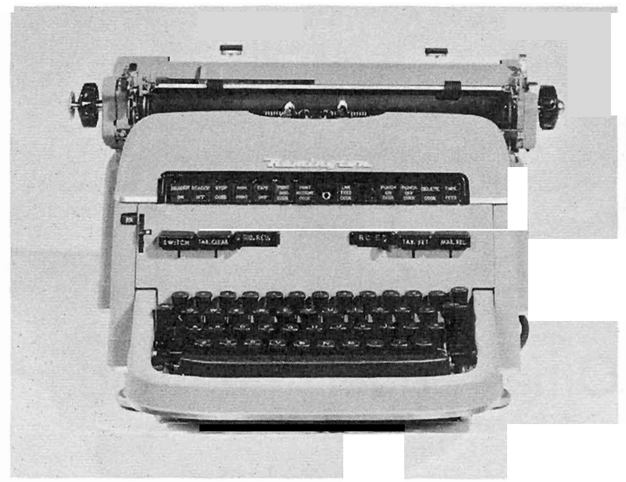


FIG. 8-2 Remington Synchro-Tape Typewriter.

Typewriter (Fig. 8-2) are the two most common paper-tape-reading-and-punching typewriters. Both of these units have the basic ability to

1. Type in the same manner as any other electric typewriter
2. Punch a corresponding paper tape while a document is being typed
3. Read a paper tape and actuate the typewriter and/or the punch in accordance with the information on the tape. (The tapes can be duplicated in this manner.)

Both of the units read and write paper tape at the rate of about 10 characters per second. The Flexowriter has been adapted so that punches on the edge of a document (see Fig. 8-1) can also be read if they follow the coding and format of the regular tape. That is, the punched edge of the document must look (to the machine) the same as a paper tape. The ability to read edge punching provides extra flexibility because information can be typed or printed on the body of the card to identify the punching or to provide auxiliary information about the punching, or the punching can represent some of the essential information written on the form or document.

Other *Equipment.* Other paper tape equipment is available to perform such tasks as: reading paper tape, punching paper tape, merging two tapes to generate a third tape, justifying lines of typing for page composing, and recording data from cash registers and other devices. One particular unit, the Friden Selectdata, permits files of paper tape to be searched for records with particular codes.

Another type of equipment such as the Robotyper and the Auto-Typist (Fig. 8-3), using a large selection of numbered records on punched tape, types the records corresponding to the button numbers pushed by the typist. This type of device is extremely useful for typing personal (i.e., "individual") letters consisting of standard paragraphs.

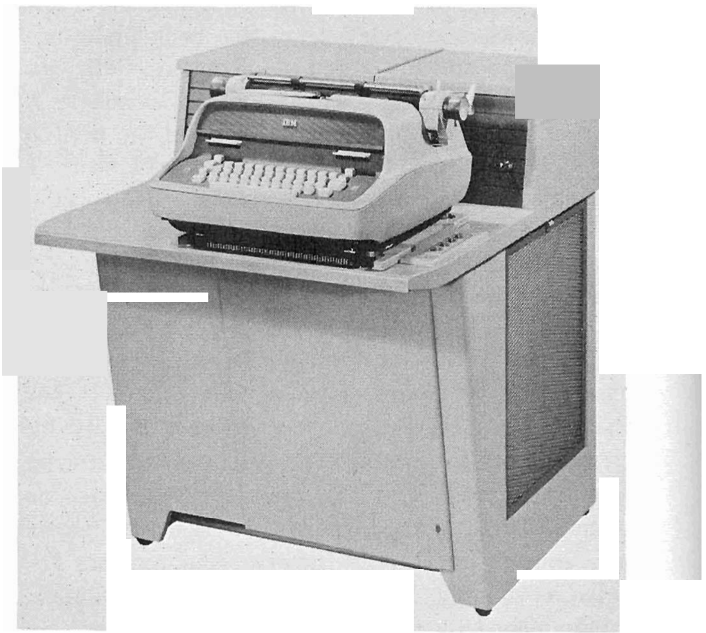


FIG. 8-3 Model 5600 Auto-Typist.

The typist pushes the button numbers corresponding to standard paragraph numbers to have the letter typed automatically (after manually entering the address). This technique is used by insurance companies' and other organizations that receive a great many inquiries that can be answered by a group of stock paragraphs. The letters are typed at the rate of 135 words per minute, and 1 clerk or typist can keep 4 machines running at essentially full speed.

Representative Costs

The costs of several models of paper tape equipment are given in Table 8-1. The costs of the paper tape and card stock will vary a great deal, depending upon the type of material and printing required. Some representative costs are given in Table 8-2.

TABLE 8-1

Representative Costs of Paper Tape Equipment

Item	Approximate Monthly Lease Cost	Approximate Purchase Price
Remington Rand Synchro-Tape Typewriter	$150	$3500
Friden Flexowriter	90 to 170	2600 to 3100
Friden Selectdata	65	2000
Friden Tape Reader	20	620
Friden Tape Punch	20	610
Friden Justowriter Set	180	5500
Friden Code Converter	110	3325
Robotyper Co. Robotyper	100	1340
Robotyper Co. Tape Perforator	35	365
American Automatic Typewriter Co. Auto-Typist		
(standard model)	—	950
(push button model)	—	1375
Royal McBee Royaltyper	100	1950
IBM Model 46 Tape-to-Card Punch	155	6680
IBM Model 47 Tape-to-Card Printing Punch	180	7630
IBM Model 63 Card-Controlled Tape Punch	110	5350

Note: The exact price will depend upon the number of accessories or optional features which are used.

TABLE 8-2

Representative Costs of Paper Tape Supplies

Item	Approximate Purchase Price
Standard paper tape	$ 30 for 30 1000-foot rolls
Heavyduty paper tape	100 for 30 1000-foot rolls
Fan-fold paper tape	2 per package of 950 feet (8½-inch folds)
Fan-fold card stock	20 for 2000 cards (3 by 7 inches each)

Applications

Utilization for Duplicate Copies. Paper tape equipment is often used to make duplicates of an original document. This is done by typing the original document and producing a punched paper tape a's a by-product. The paper tape is then run through the machine at any later time to produce another original document. One (first copy) document is produced for every pass of the tape. There are more efficient ways to make additional copies (e.g., carbon paper and other office copying equipment), but the paper tape method makes each copy look like an original personally typed document. This method is often used to send a large number of identical "personal" form letters (e.g., questionnaires, invitations, direct mail advertisements) at a relatively small cost. A typist is needed to type the address and salutation, while the remainder of the letter is produced by the punched paper tape. As mentioned earlier, a collection of standard paragraphs for letter writing can also be maintained on tape for use in composing and typing correspondence.

Multiple copies of library cards or file cards can be prepared simply with paper tape. Many organizations currently do this by typing each card separately or by preparing a duplicating master. Paper tape equipment can be used for this operation by generating a tape during the typing of the first card, then using the tape, with the ends connected to form a loop, to produce as many additional copies as desired.[1] The card stock can be a continuous fan-fold paper form which is separated after typing. This eliminates the card insertion time and pre-

[1] Turner, L. D., and J. H. Kennedy, ***System of Automatic Processing and Indexing of Reports,*** Report UCRL-6510 of the University of California Lawrence Radiation Laboratory, Livermore, California, pp. 46 (July 1961).

vents the top line of the card from being smeared during typing. As a by-product, the tapes used to prepare the cards can also be used to prepare the duplicating masters for the accession list or bulletin of new acquisitions. The tapes could also be filed for later use in making bibliographical lists and additional catalog cards. The paper tape can be filed on spools, in small cardboard cartons, or fan-folded in envelopes or card files.

Utilization for Subsequent Printing. For an order-invoicing operation in a business that has fairly permanent customers, pre-punched and filed paper tape can be used to record such things as the customer's full name, address, discount rate, terms of payment, salesman number, sales territory, and preferred method of shipment. This fixed information, which must be written on an invoice every time a customer re-orders, can be put there by reading the paper tape into the typewriter. This is an example of a situation in which a block of semi-permanent information must be added at many different times to a particular type of document. Some economy may be achieved by using a pre-punched paper tape from a file of customer tapes to prepare the fixed portion of the invoice. The variable data such as the date and the details of the order can be entered manually. If there is a standard product line and product price, this information may also be entered on the invoice from a file of product tapes. For library work, references or abstracts could be punched into paper tape, for the subsequent preparation of bibliographic lists of selected items in response to an inquirer's request.

Paper tape can be used to advantage when a number of similar documents or duplicating masters must be typed every day (e.g., menus, production schedules, or library routing lists). A tape is generated for the information that is to be repeated on each mat (e.g., the form title, column headings, and other fixed data). The tape is read into the typewriter to automatically type the fixed data, stopping automatically to permit the typist to enter the variable data.

Journal citations, part descriptions, names, chemical compounds, or other data on paper tape strips could be assembled and organized for typing to simplify editing and composing for publication. Unit-records in paper tape form can be used to compose material for publication in much the same manner as punched tab cards were used with tabulating and photocomposing equipment. The Armed Services Technical Information Agency (ASTIA) presently uses Synchro-Tape Typewriters to prepare separate tapes for each document entry in their Technical Abstract Bulletin (TAB). These tapes are then assembled and sequenced to print the final reproduction copy for each periodic TAB issue. These same tapes are then used as input to the computer equipment to add the new accessions and their indexing information to the master file for literature searches.[2]

Utilization for Input to Other Machine-Languaye Systems. When a teletype message is to be sent from one place to another, the message is often typed initially on a standard form and given to the teletype operator, who then keys the message on the teletype machine keyboard (duplicating the keystrokes of the original typist). Thus the message is keyboarded twice. This second keyboarding operation and its attendant error possibilities may be eliminated by generating a paper tape when the message is first typed, then using a paper tape reader for mechanical input to the teletype machine. This removes the necessity for the second manual keyboarding, reduces the possibility of error, and increases the productivity of the teletype machine and operator. In this example, the paper tape, produced as a by-product of an earlier operation, can be utilized for later input to a machine system. ASTIA provides another good example of such use in preparing paper tapes at the time that the Technical Abstract Bulletin copy is being typed and using these tapes to read this information into the computer for subsequent file-searching operations.

Paper tape punches are often used for point-of-action transaction recording, such as the recording of the details of each transaction at the cash registers in stores and markets. In this type of operation, the tapes accumulate the transaction data, and are collected at periodic intervals (e.g., daily) to be read directly into a data processing system for sales analysis and inventory control. Tapes which are generated while typing new or revised insurance policy

[2] Dunlop, W. W., *Automation of ASTZA—1960,* Armed Services Technical Information Agency, Arlington, Virginia, AD-247 000 (December 1960).

Hammond, W., *Evolution of the ASTIA Automated Search and Retrieval System,* Armed Services Technical Information Agency, Arlington, Virginia, AD-252 000 (January 1961).

statements can also be used as input to data processing systems for the updating of the policy records. The equipment might be used in a similar manner in libraries for order and cataloging operations.

Punched paper tape can be used as an input or storage media for computer systems if the computer equipment has the necessary paper tape equipment and a compatible code and format. Some computer equipment cannot read paper tape at all, and most computer systems will have difficulty reading the paper tape prepared for another manufacturer's computer.

Utilization for File Searching. In a few instances, paper tape has been used as the storage medium for an information-searching system, as in the Western Reserve University Searching Selector (see Fig. 8-4). Despite the mentioned limitations of paper tape searching equipment, the WRU Searching Selector was relatively successful in its operating lifetime, primarily because it was fundamentally a research and teaching tool, with some utilization as a low-volume production searching device. Much useful information was obtained from the development and application of this unit, and it is mentioned here primarily because of its historical value.

Coded document abstracts were punched into paper tape to form a machine-language index to a collection of documents. The index actually consisted then of rolls of punched paper tape. File searching was accomplished by scanning or reading the entire collection of tape under the control of a central logic control unit which was programed with specific search criteria.[3]

Paper tape devices are generally unsuitable for practical searching applications for five main reasons: (1) the reading or scanning rate is usually very slow (nominally 10 characters per second for most machines, and very seldom higher than 2000 characters per second); (2) the paper tape wears out and must frequently be recopied, especially for high-speed handling; (3) it is awkward to make changes or revisions in the tape; (4) for many operations, the serial tape is not as convenient as unit-record media like tabulating cards; (5) the entire portion of the file selected for detailed searching must be scanned in serial fashion with no possibility of reading any of the information in parallel to speed up the process. The paper tape systems do have the distinct advantage that the input-output equipment (i.e., the punch and reader) is much less expensive than the input-output equipment associated with any other storage medium.

[3] Perry, J. W., "The Western Reserve University Searching Selector," Chapt. 18 in *Tools for Machine Literature Searching*, J. W. Perry and A. Kent, editors (Interscience Publishers, New York, 1958).

Melton, J. L., "Searching the Metallurgical Literature for A.S.M.—Programming the WRU Searching Selector," Chapt. 24 in *Information Systems in Documentation*, J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

Kent, A., J. Melton, and C. Flagg, "Abstracting, Coding, and Searching the Metallurgical Literature for A.S.M.—the WRU Searching Selector," Chapt. 23 in *Information Systems in Documentation*, J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

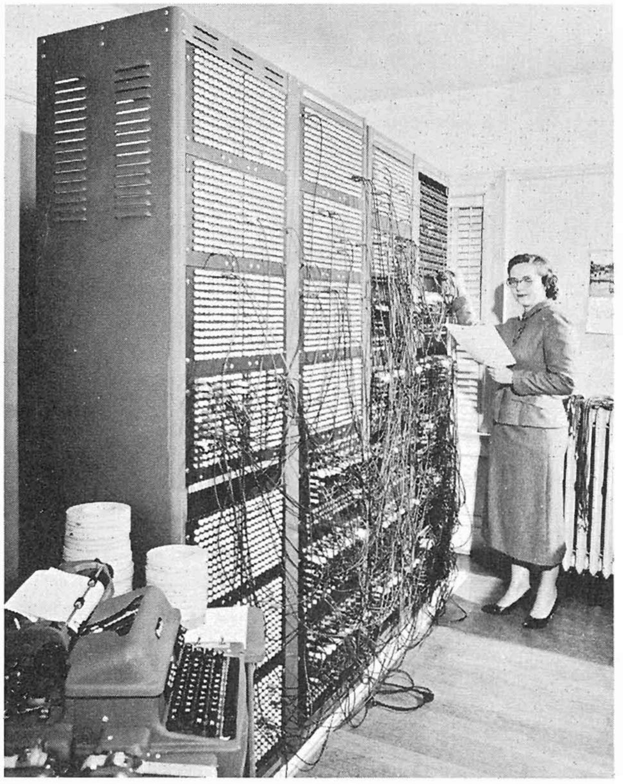


FIG. 8-4 Western Reserve University Searching Selector.

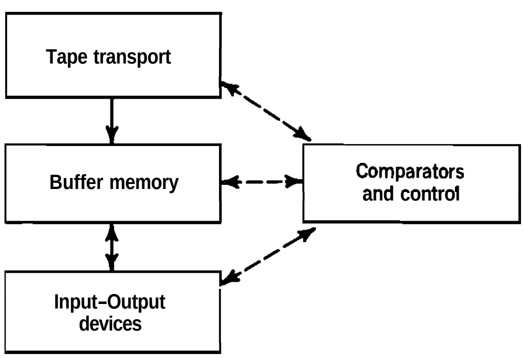
MAGNETIC MEDIA EQUIPMENT

The Equipment

All of the magnetic tape and magnetic card systems store information in some form on a magnetic material. A mechanical transport system is used to provide some relative motion between the magnetic material and the reading and writing heads. Electronic circuitry is also required to read, write, and interpret the signal, perform the control operations, and power the tape-handling mechanisms. The details of the recording and coding of magnetic tape were given in Chapt. 4 and tape transport equipment was shown in Chapt. 7.

The principal advantages of magnetic tape are three: **(1)** a very large amount of information can be stored in a relatively small volume; **(2)** a rather rapid information transfer rate to or from the tape can be maintained; **(3)** the storage is permanent so that the tape can be used many times before there is a possibility of the introduction of an error. The principal disadvantages of magnetic tape are two: **(1)** the information is usually recorded in a continuous stream in such a way that it is awkward or difficult to insert new information; **(2)** except for very special cases, the information on tape cannot be seen or read directly by humans—consequently, electronic equipment is necessary for the inspection of any information on the tape.

All of the magnetic media searching systems have common features, and might be described by the general block diagram shown below.



The tape transport or its equivalent is common to all systems, and can be any of the commercially available digital transports. Some mechanism, such as an electric typewriter, paper tape reader, or punched card reader, is usually used to record information on the tape, although the tapes could be generated by computer equipment. Some mechanism, possibly a plugboard or any input device, is used to present the questions to the system. An output device, such as an electric typewriter, paper tape punch, card punch, or high-speed printer, is usually used to record the search results. Some central electronic control and buffering system must also be used to coordinate the equipment and perform the required logical operations. The capabilities of the tape searchers differ from those of computers primarily in the fact that the tape systems cannot perform any arithmetic operations and are usually not controlled by a stored program.

Any arbitrary tape format could be used. However, if the machine could work with formats which were the same as that used with a large number of computers, then it would be possible to search tapes initially prepared on these computers. This would provide some very definite advantages such as the possibility of distributing tapes to a number of searching systems from a central computer system. In this way, a central information center could continuously distribute up-to-date listings or indexes to many satellite or auxiliary information centers. The use of a central facility to prepare distribution tapes might also reduce the local searcher equipment complexity by eliminating the need for a capability to continuously maintain the tape file and keep it up to date. The local searching equipment may require the capability to augment its assigned file with information of interest primarily to the local system. Some information services have already started a "periodic publication" of magnetic tapes in this manner.

Most of the jobs that can be done on a tape searcher can also be done with a computer, and the computer can also do many other jobs besides. For this reason, the tape searchers can hardly be justified if they are more expensive to use than the available computer equipment. Most of the computer equipment is too expensive to use solely for this task. However, several of the smaller computer systems commercially available would appear to be competitive in both performance and cost with some of the tape searchers so-far developed. As with any of the mechanized retrieval systems, tape searchers are effective only when the information they handle can be indexed or described in a rather exact manner. And as with computer equipment, the tape searchers can be used to advantage when a large number of searches must be performed in a relatively short amount of time, or when the searches are too complex to be conveniently handled manually, or when the file is too large for effective manual operation.

One problem that faces the user of a tape searcher is the initial generation and subsequent maintenance of the magnetic tape file. The file and indexing information must first be put into machine-readable form, with some control or editing to en-

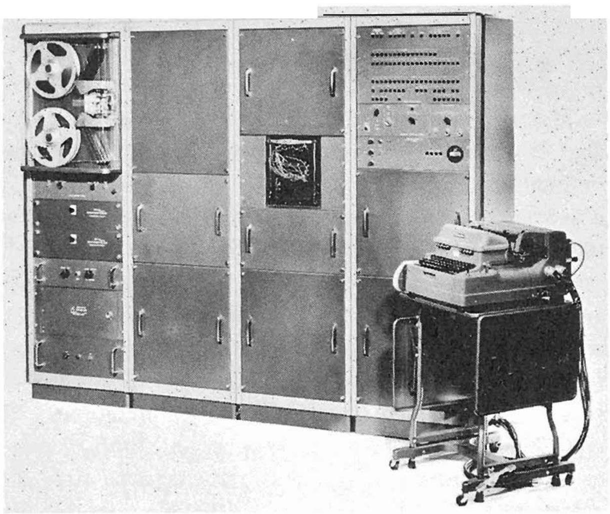


FIG. 8-5 CCC Index Searcher.

sure the validity of the transcribed information. After the master tape has been generated, there will probably be a continuing need to correct or delete items on the tape, as well as to add new information. Unless an additional tape transport is available, file additions will probably have to be placed at the end of the file instead of being inserted in order in the middle of the file. If the file does not require any particular sequencing (e.g., files with superimposed coding), then this does not become a problem. The following sections provide a brief description of the characteristics of several of the magnetic media searching systems developed to date.

The CCC Index Searcher. The CCC Index Searcher (see Fig. 8-5) was the first high-speed magnetic tape searching system developed primarily for document index searching. Sponsored by the U.S. Air Force as part of a coordinated research program in machine aids for information retrieval, it was built in 1958 by Computer Control Co., and designated as the Document Data Index Set AN/GSQ-26.[4]

The Index Searcher uses 1-inch magnetic tape in 2400-foot reels, so packed that the indexing data for approximately 70,000 documents can be stored on 1 reel. The source information is loaded on the magnetic tape from punched paper tape. The search words (total of 140 characters) are entered into the buffer memory from paper tape. The logical relationships between the search words are specified by plugboard wiring and there is enough flexibility in the specification of the search logic that several questions can be asked simultaneously. A tape-editing mode is incorporated, permitting the deletion of specific documents from the tape or the printing of the entire contents of chosen documents. A "regenerate" mode of operation is also available to copy one magnetic tape onto another, for duplicate or updated files.

The entire collection of 70,000 documents can be searched in approximately 4.5 minutes. If a document satisfies the search question, the document number and other identifying information, as well as the question number, are printed out. The basic system will (1) accept input data and write them on magnetic tape, (2) copy one magnetic tape onto another, (3) search the tape and print the selected document number and two words of additional data as well as the question number, (4) search the tape and print the entire indexing entry for the selected documents, and (5) delete specified index entries from the tape.

General Electric GE-250 Information Searching Selector. This unit (see Fig. 8-6) was developed by the Computer Department of the General Electric

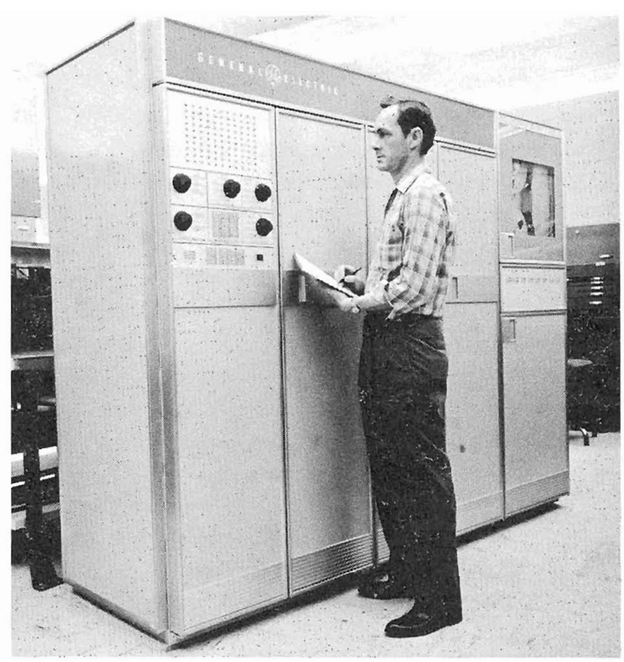


FIG. 8-6 General Electric GE-250 Information Searching Selector.

[4] Kessel, B., and A. De Lucia, "A Specialized Library Index Search Computer," ***Proceedings*** of ***the 1969 Western Joint Computer Conference,*** pp. *57-59* (Institute of Radio Engineers, New York).

Company.[5] The first unit was built and scheduled for delivery in 1960 to Western Reserve University for use with operational information services (e.g., to be used to conduct searches of the world's metallurgical literature for interested members of the American Society for Metals), as well as for use as a research tool in general studies of information systems. The equipment was removed from the company's product line and no deliveries were made; it is of interest here because of its historical significance. It was supplanted by the GE-225 computer with a special searching program.

The GE-250 logic reflected many of the techniques used with the paper tape Searching Selector at Western Reserve University. The system philosophy followed the telegraphic abstracting and coding system used at Western Reserve, although it does appear that the user could apply other coding techniques. The magnetic tapes were compatible with the low-speed IBM tapes so that they could be generated or maintained with most of the IBM computers. The basic file record consisted of 300 alphanumeric characters, which might typically be divided into a document number, bibliographic data, and an abstract. A 2400- or 3600-foot reel of tape was used, which provided storage for 4,000,000 or 6,000,000 characters, or approximately 13,000 or 19,000 documents, respectively. File searching was done at the rate of 6000 characters per second. A complete search of a single reel of tape would take at least 16 or 24 minutes for 2400- and 3600-foot reels, respectively. As many as 10 searches could be conducted simultaneously.

A large variety of input-output and optional equipment was planned, such as high-speed printers, additional tape transports, and high-speed card readers, with the first system scheduled to use a Flexowriter for input-output. The search question was to be entered into the system by typing the question statement with its logical connections and brackets on an input Flexowriter or reading a prepunched paper tape on the Flexowriter. No plugboards were used in the system. The basic system would (1) accept input data and write them on magnetic tape, (2) search the tape and print the search number and the selected document number, (3) search the tape and print the search number and selected bibliographic data, (4) search the tape and print out the complete record, (5) print out the search question.

The use of a punched paper tape for entering the search question provides a very cheap and convenient form of storage for questions which must be repeated at frequent intervals. This is exactly the situation that would arise with a machine which was used for "current awareness" or "press clipping" services to periodically interrogate the file input with standing questions. In addition, paper tape can be prepared at a time and location which do not interfere with the operation of the machine. It would even be possible to prepare, and mail or telegraph the search tapes from many remote sites directly to a central search facility.

General Electric Search Comparator. The GE Search Comparator, developed in 1961, is a smaller version of the GE-250 described earlier. This unit also uses computer compatible magnetic tape, and makes provision for a number of optional input-output devices such as an electric typewriter, punched paper tape, and punched cards. The basic file item normally consists of an alphanumeric record such as an English abstract or citation. Extensive logical capabilities are available for the file-searching or statistical analysis operations, and the search result may be a listing of selected item numbers or an expanded alphanumeric description such as a citation. The file can be searched with several questions simultaneously, and the tape may be searched while it is running forward or backward—thus eliminating the tape rewind time. The search of 900,000 words on a single reel of tape can be made in less than 5 minutes.

Heatwole *H-44*. The H-44 is a relatively simple and inexpensive magnetic tape searching system developed by Heatwole Associates (see Fig. 8-7). This equipment records in its own unique format on conventional audio recording tape, and uses relatively simple keyboard devices, switches, and visual tube displays for input-output. The basic file item on tape consists of a document number represented by zero to eight digits, and any desired number of subject headings represented by three-digit, four-digit, or five-digit codes. Using five-digit document numbers, and with an average of 4 four-digit codes per document, about 37,500 documents can be represented on a single reel of tape.

Search prescriptions are entered into the machine

[5] Gull, C. D., and P. O. Dodge, "The GE-250 Information Searching Selector," Chapt. 21 in *Information Retrieval and Machine Translation,* Part 1, A. Kent, editor (Interscience Publishers, New York, 1960).

by setting various selector switches on the console. The search is restricted to 4 subject codes at a time in all desired combinations of the four (logical sums, logical products, combinations of sums and products, and negation). All codes that are to be searched at the same time must have the same number of digits. For this single-search prescription of 4 descriptors, the file will be searched at the rate of approximately 1000 documents per minute. Since a serial scan is made of the entire tape, the order or sequence of the documents on the tape is not critical. New material can continually be added to the end of the existing tape.

The result of a search can be either a list of the selected document numbers or a count of the number of selected documents. The counting feature permits the equipment to be used for statistical analysis of the file contents.

Magnavox Magnacard. The Magnavox Company, with financial support from the U.S. Air Force, has developed equipment and techniques for high-speed handling of magnetic cards (hence the name Magnacard) to perform such functions as sorting, collating, and file searching. This system has the inherent flexibility of the unit-record systems, while also providing the high transfer rates and storage capacity associated with the magnetic recording techniques.[6]

[6] Hayes, R. M., "Magnavox Activities in Data Processing," a section in *Modern Trends in Documentation,* M. Boaz, editor, pp. 74–81 (Pergamon Press, New York, 1959).

Hayes, R. M., "The Magnacard System," Chapt. 19 in *Information Retrieval and Machine Translation,* Part 1, A. Kent, editor (Interscience Publishers, New York, 1960).

Hayes, R. M., and J. Wiener, "Magnacard—A New Concept in Data Handling," *1957 IRE WESCON Convention Record,* Part 4, pp. 205–209.

Nelson, A. M., H. M. Stern, and L. R. Wilson, "Magnacard—Mechanical Handling Techniques," *1957 IRE WESCON Convention Record,* Part 4, pp. 210–213.

Burkig, J., and L. E. Justice, "Magnacard—Magnetic Recording Studies," *1957 IRE WESCON Convention Record,* Part 4, pp. 214–217.

Nelson, A. M., *The Research and Development of the Magnacard System,* Wright Air Development Center Technical Report 58-421, ASTIA Document No. AD-211 694 (September 1958).

Wilson, L. R., *Integrated Card Handling and File Unit,* Magnavox Research Lab. Report R-332, Rome Air Development Center Report No. RADC-TR-60-167, AD-243 065 (May 1960).

Westgard, R. E., "The Magnacard System," *Datamation,* Vol. 7, No. 7, pp. 42–43 (July 1961).

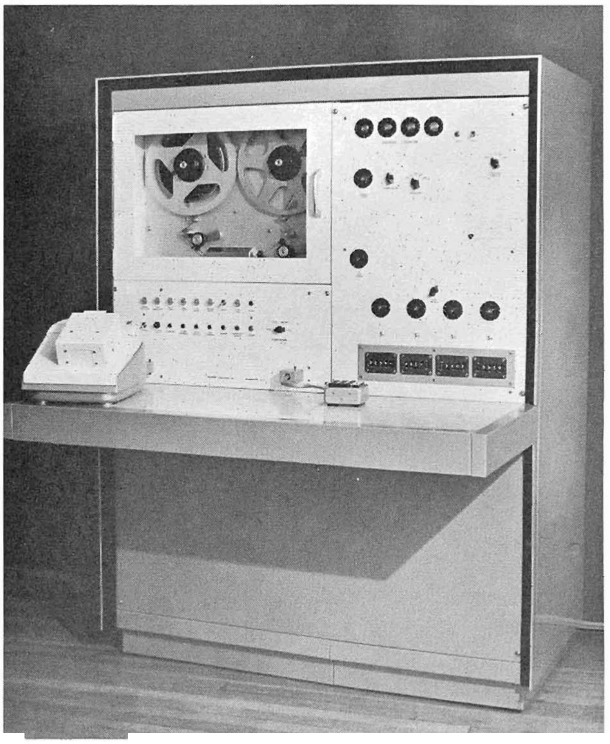


FIG. 8-7 Heatwole H-44.

A laminated 1-by-3-inch card is used, with a Mylar base, an iron oxide coating for the recording material, and a Mylar overlay to protect the oxide coat during card handling. Each of these cards has a capacity of 756 alphanumeric characters, which is a great deal more than the punched card unit-record. With improved techniques, it should not be too difficult to increase this storage capacity even further. The cards have an estimated operating life of 20,000 cycles through the operating equipment. The cards are stored in drawers or magazines, each of which can hold 3000 cards. The drawers can be automatically moved by program control to and from the card-handling equipment, and the cards loaded or unloaded from them. Ten or more drawers may be assembled together to form a file block: A 10-drawer file can store a maximum of 30,000 cards.

The system philosophy is to perform the sorting, selecting, and merging functions by physically moving the record instead of reading it and copying the information onto a record to be stored at a different location. The card transport mechanisms are pri-

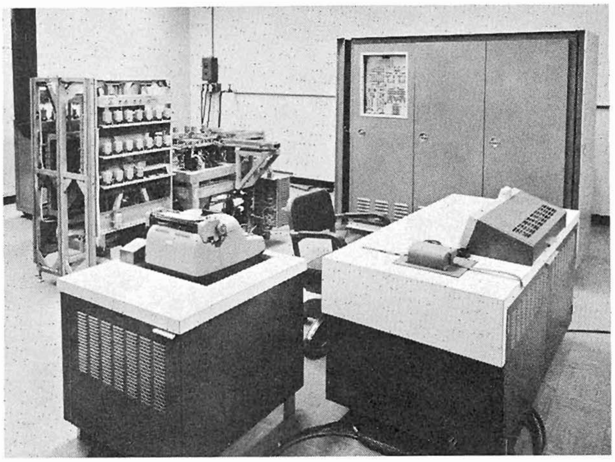


FIG. 8-8 Magnavox Magnacard equipment.

marily pneumatic devices: vacuum transport drums, card-feeding and card-stacking stations, drum-to-drum transfer devices, and card-holding devices. An illustration of a complex of Magnacard equipment is shown in Fig. 8-8. The cards are picked out of the trays and transported on the drums at a peripheral surface speed of 300 inches per second. By appropriate plugboard and pneumatic controls, the cards are transported to the selected receiving drawer. It is possible to pass the cards from one drawer to another at rates of up to 90 cards per second.

Although the Magnacard system can be used in a serial manner in the same way as other tape searchers, its real benefit comes when the user takes advantage of its unit-record characteristics. In this way, information can be added to or deleted from the file without rewriting the file contents. In addition it should be possible to develop an entire family of equipment similar to the tab equipment, so that the Magnacard can be used as an extension of the conventional punched card techniques. One of the real advantages of unit-record systems which might be further developed here is their ability to be produced or created at many points, and later assembled and processed.

One other potential use of the Magnacard concept is for an image storage and retrieval system, that is, to store an image instead of completely digital data on the card. The image could be a video image as in television recording systems, but it would more likely be a conventional film image with an accompanying magnetic coating to carry the descriptive information in digital form. This latter approach has been studied for possible applications to image storage and display systems.[7]

RCA Video File. RCA has proposed an image retrieval system based on the storage and retrieval of video images on 2-inch-wide reels of magnetic tape in much the same manner as television video signals are recorded. Original documents would be filmed with a TV-type camera and stored on tape in the same manner as recorded TV programs. Additional indexing information or document numbers would accompany the image in digital coding, so that the file contents could be searched digitally in much the same manner as with the tape searchers mentioned previously. The results of a search (i.e., the selected images) would be displayed on remote TV monitors for viewing or photographing, or printed on equipment such as the A. B. Dick Co. Videograph which could provide page-size reproductions of the images at a rate of about 1 to 2 pages per second. The storage density would be rather high, with copies of approximately 36,000 page-size documents stored on a single 7200-foot roll of tape. A high-speed search would pass over this roll in about 5 minutes. The original documents could be scanned and recorded on tape at a rate of about 100 sheets per minute.

Some of the general problems facing the user of video storage equipment are: (1) the recording system will probably not provide high-quality images; (2) the images are stored in an analog manner, and are subject to some degradation during every copy or regeneration process; (3) the performance of video or TV tape recorders is currently such that a tape can only be used a few thousand times (maybe as high as 20,000 in some cases) before it becomes too scratched and worn for decent service. The combination of these last two considerations suggests that there is no such thing as permanent or archival storage of the image. The original recording can only be searched a few thousand times before a new tape has to be regenerated. But the regeneration process degrades the image, so that in time, the image has been degraded to a point where it is no longer of use. These two problems also make it difficult to modify or update the master

[7] Gelb, J., *Utilization of Magnacard as a Display Device*, Magnavox Research Laboratories Report R-357, Rome Air Development Center Report RADC-TR-61-1, AD-251 271 (February 1961).

file. It would appear that this type of system will find its best application in situations which concern themselves with relatively short-term document lives instead of situations that require permanent or semi-permanent storage.

Representative Costs

Most of the systems described in this chapter have been essentially "one-of-a-kind" developments, and their manufacturers would probably be willing to modify any future equipment to suit the needs of a particular user. Since there are no fixed product specifications, exact prices cannot be quoted for all of these systems. However, some rough estimates can be made of the costs of most of these basic systems, and these are given in the following table:

System	Approximate Purchase Cost (dollars)
RCA Video File	½ to 1½ million
Magnavox Magnacard	over 100,000
CCC Index Searcher	100,000
GE-250	100,000
GE Search Comparator	40,000
Heatwole H-44	10,000

In comparison with the capital equipment cost, the remaining variable costs (supplies, maintenance, power, etc.) are relatively minor. For example, digital magnetic tape may be obtained for 70 to 100 dollars per reel, magnetic cards for 30 to 100 dollars per thousand, and video tape for about 425 dollars per roll. At this time, not enough operating experience has been obtained with these systems to accurately describe the relative cost factors.

Applications

To date, the use of all of this equipment has been restricted to experimental situations. None of this equipment has been used in a regular production application.

Additional References

James, P., "A Photo-Magnetic System for Document and Information Retrieval," American Documentation, Vol. 10, No. 4, pp. 286–295 (October 1959).

Mooers, C., "The 'Tape Typewriter Plan,' A Method for Cooperation in Documentation," Association of *Special* Libraries and *Information* Bureaux Proceedings, Vol. 12, No. 8, pp. 277–291, August 1960.

Basic Microforms
Mechanized Micro-Image Systems
Applications
Costs

9

Microfilm and image handling equipment

Microphotographic techniques were first developed around 1839, and have been in use since that time.[1] However, until a few years ago, microfilm was used primarily for storage purposes—for reducing the volume of filing space required for the storage of inactive documents, or providing a security copy in case the original was destroyed. Microfilm techniques, for example, were used during World War II to reduce the tonnage of letter mail shipped overseas. In business it is often necessary to save many original documents--or copies of original document—for long periods of time, because of legal or accounting requirements. In order to reduce the bulk of this volume of records, many companies have instituted the practice of making photographic reductions of the original documents, and then destroying the original documents. Often, microfilm copies of all the records have been made and then stored at a location separate from the originals as a precaution against the loss of the original documents by fire or other circumstance. In nearly all such cases, the film copies were essentially inactive files that required very infrequent access, and the microfilm systems and procedures were therefore rarely organized for quick and convenient searching.

The early microfilm work gave little consideration to the use or development of high-resolution camera and film systems although some of the first work by Dancer in 1839 produced a microphotograph at a reduction ratio of 160:1. Film copies were poorer than the copies that can be made now. This is probably due in large part to the fact that the filming of business records and legal documents did not require the same degree of detail as the filming of engineering drawings and maps currently requires.

Some organizations do their own microfilming, especially if they have a large volume of work or if the work does not have to be done with great precision. Relatively few, however, develop or process their own film. Many microfilm users prefer to contract the work to microfilm service bureaus, which specialize in microfilming operations. Arrangements are often made for a microfilm crew from a service bureau to set up their camera equipment on location in a client's office and convert a firm's records to microfilm form in a very short time. However, as more easily operated equipment becomes available, more organizations will find it economical to do their own microfilming. Currently available equipment that is easy to use and inexpensive is adequate for the filming of business records. However, the recording of very detailed material, such as maps and engineering drawings, will continue to require skilled operators and an extensive amount of equipment.

Film sizes (referring to the film width) of 8mm (8 millimeters), 16mm, 35mm, 70mm, and 105mm are the most common. The 8mm and 16mm sizes (see Fig. 9-1) are used primarily for business-record copying and would be difficult to use for work requiring precision. Detailed maps or drawings are usually filmed on 35mm film or larger (see Fig. 9-2). Most of the equipment and most of the development

[1] Luther, F., ***Microfilm: A History, 1839–1900*** (National Microfilm Association, Annapolis, Maryland, 1959).

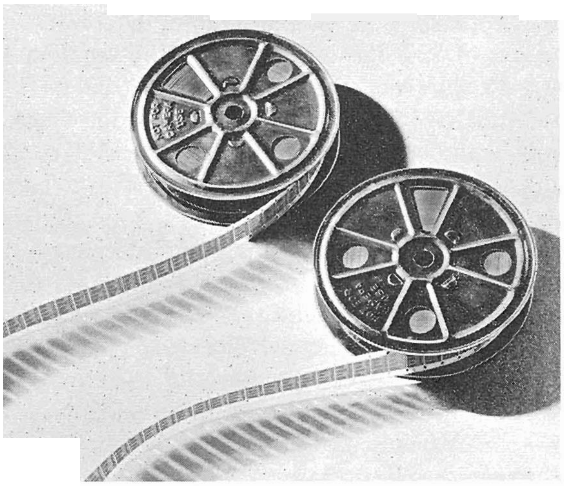


FIG. 9-1 16mm roll microfilm.

effort for high-quality microfilm systems incorporates 35mm film.

Some film equipment utilizes reduction ratios up to 200:1 from the original material.[2] That is, an original document with a 10-inch length will be recorded with a length of $\frac{1}{20}$ inch on the film. A reduction of the edge length by 200:1 will also bring a corresponding area reduction of 40,000:1. When the reduction goes beyond 20:1, such things as basic film resolution, proper film exposure and processing, freedom from camera vibration, and avoidance of dust and scratches become increasingly critical. In addition, the optical systems become increasingly complex and expensive as the reduction ratios get larger. Reduction ratios of more than 30:1 are easily achieved while recording from the original under laboratory conditions, but under the environment of a regular production situation, the reduction ratio is usually held to about 20:1 for each recording operation. The 70mm and 105mm films have been used to minimize the need for large reduction ratios. In addition, the larger film sizes can easily be used as unit-records, and handled separately from each other, without requiring an extensive amount of effort for film cutting and mounting into individual film cards or jackets.

[2] Tauber, A. S., and W. C. Myers, "Photochromic Micro-Images: A Key to Practical Microdocument Storage and Dissemination," *American Documentation,* Vol. 13, No. 4, pp. 403-409 (October 1962).

Carlson, C. O., et al., "The Photochromic Microimage Memory," in *Large-Capacity Memory Techniques for Computing Systems, M. C.* Yovits, editor (Macmillan Co., New York, 1962).

Most mechanized micro-image systems utilize 35mm film in one form or another.

For microfilm or any image storage system, one important performance measure is its ability to faithfully reproduce or play back the original image to full size after it has been recorded in microform (i.e., to produce a sharp image from the recording). This is somewhat analogous to the "fidelity" of a record player or tape recorder. Since there is some expense involved in achieving high fidelity—in optical as well as audio systems—compromises are usually made between cost and system performance. The unaided human eye can resolve or distinguish between parallel lines which are packed together at a density of approximately 16 lines per millimeter (mm).[3] This density is the normal limit to which

[3] The separability of parallel lines is one criterion of resolution. Another common criterion is the ability to sense the direction of lines which may be displayed on test charts at right angles to each other. See F. E. Washer and I. C. Gardner, *Method for Determining the Resolving Power of Photographic Lenses,* National Bureau of Standards Circular 533 (U.S. Government Printing Office, Washington, D.C., May 20, 1953).

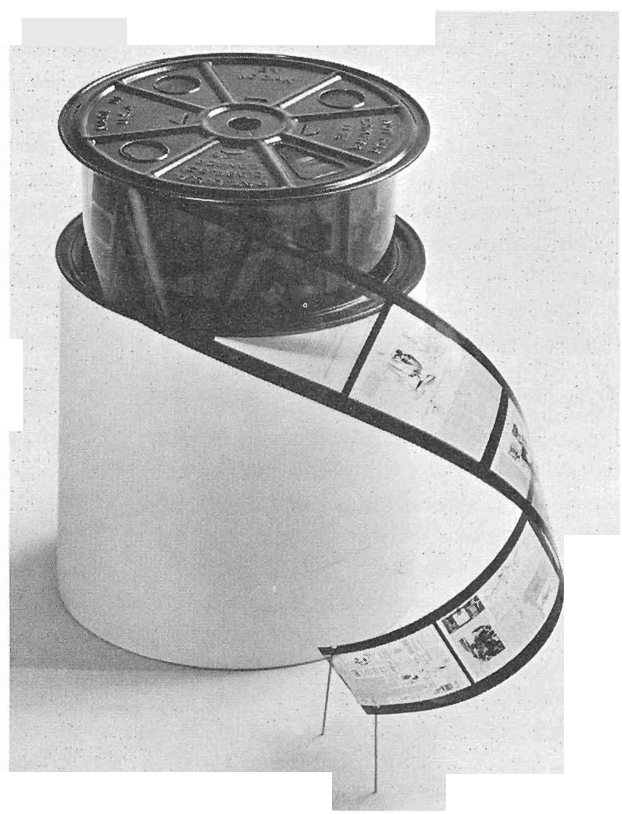


FIG. 9-2 35mm roll microfilm.

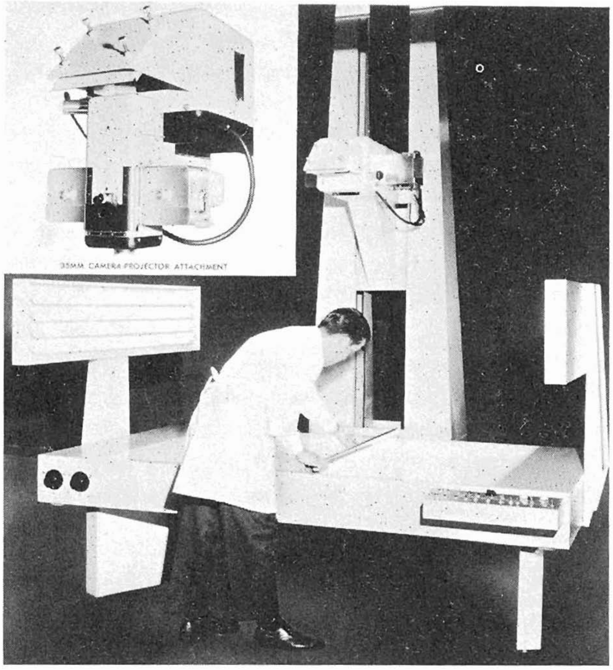


FIG. 9-3 Micro-Master Planetary camera (Keuffel & Esser Co.).

the unaided eye can separate the individual lines. Packing densities higher than this appear to be only a blurred or continuous smear. Therefore, 16 lines per mm is the greatest detail that the human eye can make use of, and could be used as the ultimate design goal for image storage systems. However, in practice, many microform equipment designers aim at blowing back from microform to a full-size resolution of about 4.5 lines per mm, which is considerably less than the limit of what the human eye would see if it were viewing the original full-size image. This end goal of 4.5 lines per mm resolution in the final image blown back from some microform determines in large part the performance characteristics required of the basic film stock and optical system. For example, to achieve this end resolution with a microfilm system operating at 10:1 reduction would require a film and optical system with a net resolution of 45 lines per mm (in order to blow the image back to 4.5 lines per mm), and a system operating at a 100:1 reduction would require a film and optical system with a net resolution of 450 lines per mm. Consequently, higher-performande film and optical systems are required if the system is to operate with high reduction ratios.

Most of the commercially available camera systems for high-quality microfilming guarantee resolutions of 120 lines per mm over the required field of view with reduction ratios of about 20:1. These high-quality cameras often achieve 150 lines per mm in the center, and decreasing resolution down to 120 lines per mm in the corners. This is because the resolving power of any camera is usually greatest at the center of the image and decreases along any radius. Such performance is adequate for the filming of most line drawings, and essentially all business record filming. The 16mm cameras for general business records normally have resolutions of about 60 to 80 lines per mm for reduction ratios of about 20:1.

The microfilm cameras are generally of two types: flatbed cameras and flow cameras. The flatbed cameras require the original document to be secured rigidly on a plane surface (horizontally or vertically) while the picture is being taken (see Fig. 9-3). The flow cameras operate with the documents placed on a continuously moving conveyor belt (see Fig. 9-4). The flow camera allows a faster production rate for the filming operation, but the movement of the document must be synchronized with the camera exposure. Some efforts have been made to improve the production rate of both these types of camera by providing devices such as automatic exposure meters and controls, but the filming is still limited primarily by document-handling speeds. The flow camera shown in Fig. 9-4

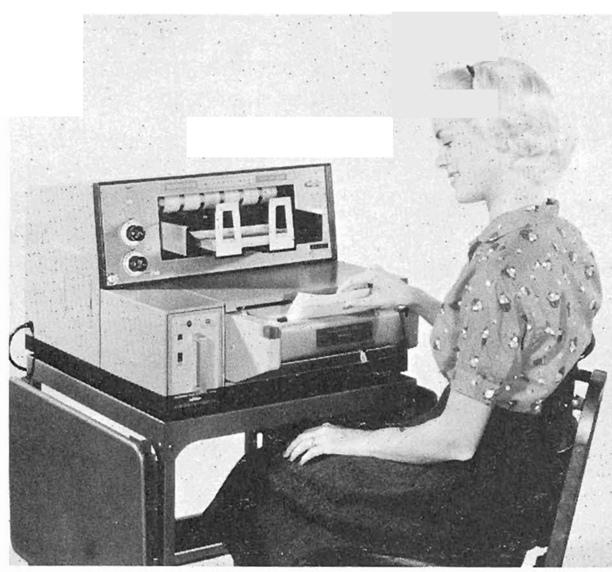


FIG. 9-4 Recordak Reliant 500 microfilmer.

is useful for filming business documents and other single-sheet material that does not require especially high-quality work. This particular 16mm camera can microfilm up to 500 check-size items or up to 185 page-size items per minute.

Film Technology

There are two principal types of commercial silver halide microfilm emulsions: (1) a general-purpose type that gives good results for most work, including engineering drawings; and (2) a type designed for recording business documents on 16mm film, where resolution can be sacrificed for higher recording speed.[4] With regard to resolution, ordinary home camera film may record about 70 lines per mm, commercial microfilm may record about 200 lines per mm, and the highest photographic resolution is obtained with Lippman emulsions or spectrographic films, which record about 1000 lines per mm.[5]

The Diazo process takes advantage of the characteristics of certain chemicals that respond to ultra-violet light by losing color or by being destroyed, and that also react with ammonia to form dyes. In operation, an ultra-violet light shines through the original subject material onto a paper or film stock coated with these chemicals. The sensitized paper beneath the dark spots on the subject document is protected from the bleaching effects of the light, but the coating beneath the clear areas of the document is destroyed. The copy stock (paper or film) is then exposed to ammonia vapors which form a permanent dye with the chemical coating that has not been bleached or destroyed. The ammonia compound can actually be used in two different forms, and these two techniques are usually referred to as a "dry" process and a "wet" process. The dry process (the familiar Ozalid process) uses ammonia vapors from ammonia water to form the dyes. The wet process actually wets the copy stock with a light coating of a special solution. The wet process does not release as many objectionable fumes as the dry process, but it is slower because the copy stock must dry. The final print for the Diazo process can be on almost any kind of stock, such as paper or Mylar, whether opaque or transparent.

The resolution of Diazo film is limited primarily by the uniformity of the surface into which the light-sensitive materials are impregnated. The finest silver-film grain size averages about 3000 Angstroms,[6] as compared to approximately 15 Angstroms for a Diazo molecule, the equivalent image particle. And whereas each generation of printing of silver halide film results in a loss of about 10 to 20 percent in resolution, there are claims that the Diazo film suffers a loss of only about 5 percent. Consequently, for a tolerable limit of resolution degradation, it would seem that several more generations of useful Diazo prints could be made than could be produced with silver film. The Diazo films, unlike silver films, are insensitive to radioactivity, both before and after processing. The Diazo image is impregnated in the film base, rather than in a surface emulsion as on silver film. Consequently, the image also suffers less than silver films from problems of scratching and finger marks.[7] The Diazo film does not burn and resists destruction by micro-organisms. Its print life is estimated to be at least 25 years, and could perhaps be several centuries under proper storage conditions.[8]

The Kalfax process uses an emulsion of photosensitive compounds on a plastic or Mylar film stock.[9] When exposed to ultra-violet light, these compounds decompose and the fragments of the decomposition set up internal stresses on the plastic. When the plastic is heated, the stresses are removed and permanent bubbles or light-scattering centers are produced. The number of light-scattering cen-

[4] Burris, W. A., "Characteristics of Silver Halide Microfilm," *Proceedings of the Tenth Annual Convention of the National Microfilm Association,* pp. 194–199 (April 1961).

[5] Camras, M., "Information Storage Density of Magnetic Recording and Other Systems," *Institute of Radio Engineers Transactions on Audio,* Vol. *AU-9,* No. 5, pp. 174–179 (September 1961).

[6] One Angstrom unit of length is equal to 0.00000001 centimeter.

[7] Kolb, F. J., Jr., and E. M. Weigel, "Protective Treatments for Microfilm," *Proceedings of the Tenth Annual Convention of the National Microfilm Association,* pp. 270–284 (April 1961).

[8] Rubin, H. E., "Sensitometry of New Diazo Films as Related to Specific Microfilm Application," *Proceedings of the Tenth Annual Convention of the National Microfilm Association,* pp. 202-225 (April 1961).

[9] Niset, R. T., "The Basis of the Kalvar System of Photography," *Proceedings of the Tenth Annual Convention of the National Microfilm Association,* pp. 177–191 (National Microfilm Association, Annapolis, Md., 1961).

Scott, P., "The Miraculous Bubble: A Look at Kalfax Microfilm," *Library Resources and Technical Services,* Vol. 3, *No.* 1, pp. 40–46 (Winter 1959).

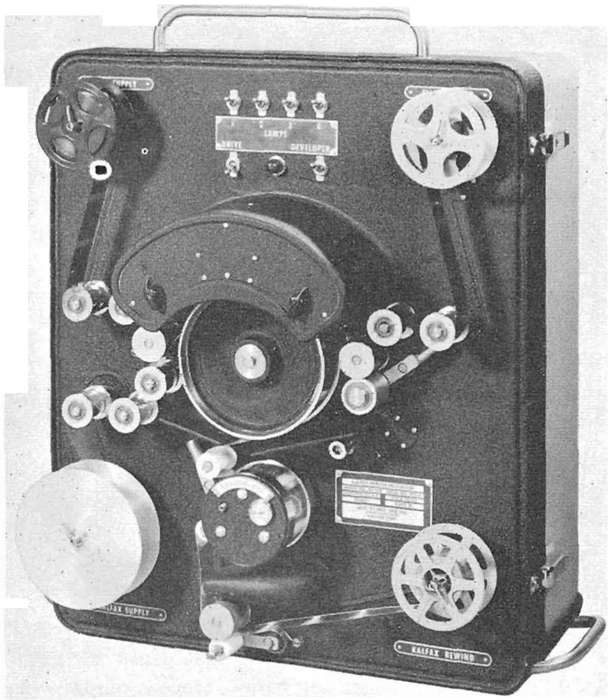


FIG. 9-5 Kalfax roll film printer-processor.

ters produced is proportional to the quantity of light absorbed. The important feature of this process is that the image is produced by exposure to light, and developed by exposure to heat. The copying process is completely dry and requires the aid of no additional chemicals. The developed and fixed film is extremely stable, and essentially unaffected by water, light, and most solvents. As with the Diazo process, no darkroom is required at any stage of the process, although the film must be developed almost immediately to save the latent image. Kalfax film, unlike the regular silver halide film compounds, is sensitive only to ultra-violet radiation and is not seriously affected by visible light or such ionizing radiations as nuclear radiation or X-rays. This insensitivity greatly simplifies the daylight handling of the film, and gives the film an exceptionally long shelf life. Kalfax film can be provided with resolutions of 80 to 120 lines per mm, which are adequate for most microfilming. Resolutions of 500 lines per mm have been obtained under controlled conditions. Image entries can be posted onto a basic Kalfax film stock record in such a manner that the basic film record or sheet can periodically have new images added to it. This feature could be used for, say, adding or updating engineering change orders to a film image of an engineering drawing, or updating public legal records such as tax records or land titles.

The Diazo technique is primarily used in microfilm form to make contact prints (transparencies) from roll microfilm. It is also used in some other forms such as Actifilm and CIM-cards, which are discussed later in this chapter. Kalfax film is useful primarily for making film copies from the original photographic film negative. It too can be used to make contact prints (transparencies) of roll microfilm. Both Kalfax and Diazo equipment are available to produce contact prints of roll and aperture card (unit-record) material, and to do this by a quick and dry process. One example of a Kalfax roll film duplicator is shown in Fig. 9-5.

Means for Viewing the Film. Most of the early microfilm systems included some provision for viewing the film—usually by projecting the image onto a surface where it could be observed. Some years later, as the viewers were improved for easier viewing, they also became equipped with provisions for making page-size copies of selected film frames. The user could simply push a button on the viewer and obtain a page-size copy of the frame in view. Figures 9-6 and 9-7 show representative viewer-printers in different sizes. Equipment is still available for viewing purposes only, using film in roll or

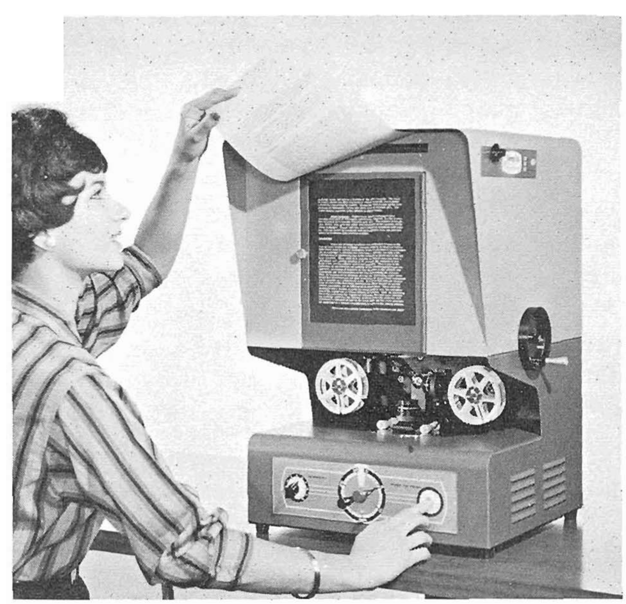


FIG. 9-6 MMM Filmac 100 microfilm viewer-printer.

cartridge form (see Fig. 9-8), but a good fraction of the viewers being purchased today are combination viewer-reproducer devices which allow a page-size copy to be made quickly and with a minimum of effort by the user. The viewing screens are available in large sizes if necessary, and the reproducing equipment for several viewers will handle aperture cards as well as roll film. Most of the newer equipment is designed for a 35mm film size.

Means for Copying the Film

Early microfilm users obtained duplicate copies of the microfilm negatives by conventional silver halide film and darkroom techniques. However, the Diazo and Kalfax films are now being employed to provide duplicate microfilm copies. A number of methods for obtaining duplicate microfilm rolls or aperture cards are shown later, in Fig. 9-42, starting with a silver negative film and ending with negative aperture cards. It should be noted that (1) Diazo film creates a negative from a negative; (2) silver film creates a positive from a negative, and vice versa; (3) Kalfax creates a positive from a negative, and vice versa. The term "negative" means that the tone values of the image have been reversed. That is, black letters now appear as white, and white backgrounds now appear as black fields. With silver film, copies of the original film negative are often made (from intermediate film positives) for use as the master, reproducible for all the necessary copy work. The original negative is thus spared the handling and rough treatment of the reproduction process while copies show relatively little loss of quality. This second roll of negative film is often cut up into separate frames and mounted in individual aperture cards to be used as the reproduction, reference, and distribution decks. The original negative is usually retained in roll form in security storage. It is common practice to send sets of drawings or other images to interested parties in the form of second- or third-generation films.

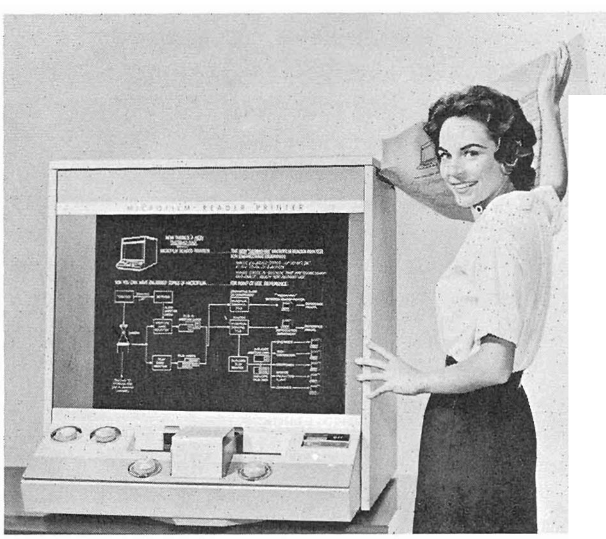


FIG. 9-7 MMM Filmac 200 microfilm viewer-printer.

FIG. 9-8 Recordak Lodestar microfilm reader.

Means for Making Hard Copy from the Film

Until a few years ago, the most common method of producing enlargements or contact prints from microfilm was to project an image onto a silver emulsion paper and process the paper in a darkroom. This produced a quality print but was expensive and slow. In the past few years, much work has been spent on devising dry copying processes such as Diazo, Kalfax, and electrostatic printing. The aim has been to eliminate the time and expense of developing, fixing, washing, and drying the film, as well as maintaining and operating a darkroom. These dry print techniques with their advantages of speed and dry processing have been developed to the point where the quality is adequate for many applications.

Silver halide techniques are still used where quality is a prime consideration. The resolution and density contrasts are greater for silver halide materials than for the other processes, and the resultant image copy usually has a greater degree of permanence than Diazo copies. The primary disadvantages are the relatively high material costs, the necessity of using a darkroom, and the requirement for the wet processing of the resultant copy. The Diazo techniques are useful, but relatively slow. The electrostatic techniques, all using essentially dry processes will probably account for most of the volume printing work for some time. All of the high-speed and high-volume microfilm printing equipment utilizes some form of the three following electrostatic processes (described by their commercial names): the Electrofax, Haloid-Xerox, and A. B. Dick Videograph processes. One Electrofax machine in use today reproduces 15 engineering drawings (17 by 22 inches) per minute from a 35mm microfilm positive roll or aperture cards. Electrofax printers in development have a planned printing speed of 120 sheets (8½ by 11 inches) per minute.

FIG. 9-10 Haloid Xerox Copyflo 1824 printer.

FIG. 9-9 Haloid Xerox Copyflo 24 continuous printer.

The Haloid Company's Copyflo Xerox equipment (see Figs. 9-9 and 9-10) can reproduce copy from microfilm intermediates such as aperture cards or roll microfilm. The final prints emerge dry and ready for use on plain unsensitized paper, vellum, or offset masters. Equipment is available that will print from microfilm at a rate of 20 linear feet per minute on 24-inch-wide paper.[10] Several microfilm service bureaus use this type of equipment to reprint, on demand, single copies of out-of-print books from roll microfilm for a charge of about 4.5 cents per page. The A. B. Dick Videograph system utilizes a special cathode ray tube to transfer the image to paper in contact with the'tube face. This process has only recently become commercially available and is potentially capable of very high

[10] Hawken, W. R., "Developments in Xerography: Copyflo, Electrostatic Prints, and O-P Books," *College and Research Libraries,* Vol. *20,* No. 2, pp. 111–117 (March 1959).

printing speeds. It is presently being used to print mailing labels for several large publishers. With the exception of the silver halide techniques, none of the previously described reproduction techniques can produce anything but a black-and-white print on today's commercially available equipment. However, development work is under way in several laboratories to develop multi-color electrostatic printing equipment.

BASIC MICROFORMS

Microfilm records are used in a variety of sizes, formats, fabrication, and indexing arrangements. There are several fundamental differences between each of the record types described in this section. For example: there are unit-records and reel records; there are records with coding for machine searching and records without specific coding; and in addition to microfilm transparencies there are also some opaque forms of micro-storage.

Microfilm Roll Storage

Conventional Systems. For the many applications which do not require frequent file reference, the bulk storage of simple roll microfilm is the least expensive of all the micro-storage methods. As a rule, very little indexing or classification of such files is necessary. Examples of this bulk storage are the microfilming of newspaper collections, customers' faces and checks at check cashing stations, old files of business documents, laboratory notebooks, and some technical documents or theses. The indexing may be accomplished by assigning a different reel number to each reel and keeping a rough record of the contents of each reel (e.g., the 1946 correspondence with Jones Lumber Company, or the November issues of the *San* Francisco Chronicle). To facilitate rapid search of a reel and locate a particular sub-section of that reel, unique block separator (or "target") symbols are often filmed after each major sub-division of images on the reel. These locator symbols are easily recognized when the frames are moving past the viewing screen at high speed. As an added aid to the viewer, the first image of the reel often contains a title and brief index to the contents of that reel. For microfilm publication in general, full bibliographic information should be included in one of the first frames to show such things as the author's name and title, date, table of contents, and index. The microfilming program should follow the general principle that the time to start the bibliographic control of the microform is when it's made. Representative equipment for viewing such film was shown earlier in Figs. 9-6 through 9-8.

Special Systems—The Addition of Coding for Machine Searching. Special equipment is available for the high-speed machine searching or selection of files of roll film. Some of these units such as the Rapid Selector, the Benson-Lehner FLIP, and the FMA Filesearch are described later in this chapter. The roll microfilm used in these systems differs from conventional roll microfilm in that a code pattern is recorded on the film to provide a machine-language representation of the indexing information for an image or series of images.

Microfilm Unit-Record Storage

Aperture Cards. The aperture card (see Figs. 9-11 and 9-12) consists of a file card with one or more frames of microfilm mounted in a window on the card. This card has been a popular record medium for many applications in which each image or small group of images represents a separate file item to be filed, retrieved, and handled as a unit (i.e., a unit-record system). The microfilm may be of any type—silver, Diazo, or Kalfax. The card stock is

FIG. 9-11 Filmsort Model 086 aperture card copier.

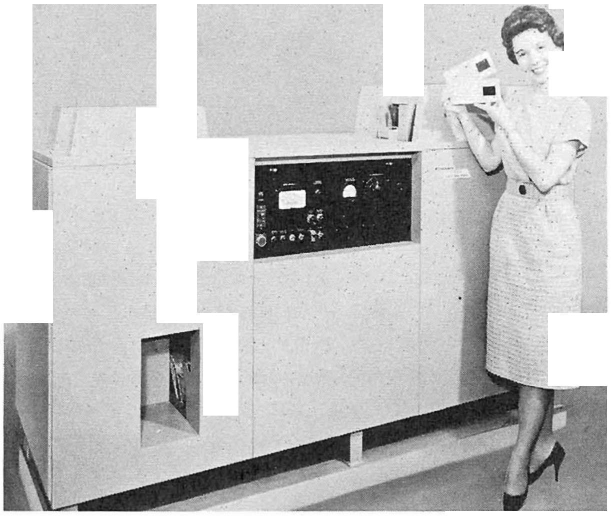


FIG. 9-12 Filmsort Model 041 automatic aperture card-to-card printer.

available in many shapes and thicknesses and with different window sizes for different film sizes. The forms may be pre-printed if desired, and some are available for edge-punched coding. However, the most popular size has the same dimensions and uses the same stock as the conventional tab card, with a single aperture for a frame of 35mm film. Three-by-five-inch card stock is also popular. The use of tab card stock allows the aperture card to be handled on conventional tabulating equipment, providing some degree of mechanization for the file system. Relatively simple manual equipment is available to cut single 35mm frames from a film roll and mount them in the aperture card. This mounting equipment may also be provided with projection and viewing facilities so that the operator mounting the cards can see whether or not his work is correct. Most of the microfilm viewing equipment can handle the aperture card as well as roll microfilm. Equipment is also available to easily reproduce the image from one aperture card onto the film of another aperture card (see Figs. 9-11 and 9-12). This card-to-card copying equipment usually copies the image onto Diazo or Kalfax film stock in the copy card. Equipment is also available for making hard copy prints (page size or engineering drawing size) from aperture cards (see Figs. 9-9 and 9-10). The aperture cards are used in many applications, although the greatest number of them are used for engineering drawing systems. They are limited to a certain record size. When the record size starts to exceed 8 pages in length, or additional data must be added at a later date, other methods such as film jackets or microtape may become preferable.

Acetate Jackets. Transparent acetate sleeves or pockets are available that will hold one or more frames of microfilm in a relatively dust-free manner. The film can be viewed without removing it from the sleeve, and only simple manual equipment is required to mount the film in the jacket. The jackets are useful when film frames must periodically be added to some master file, such as a hospital patient's records. The jackets can also be used to store all of the images of a thesis or report by snipping the original roll film into appropriate-length strips. Some jackets hold as many as 120 frames of roll microfilm. Others are attached to heavy paper stock on which descriptive information can be typed or printed. The paper stock could also be used for edge-punched coding of the jackets. Jackets are available in many film and card sizes. In their present form these records cannot be manipulated by mechanical equipment in the same manner as aperture cards. The jackets are useful as long as you need records only in one location. It becomes rather awkward and expensive to develop and maintain duplicate jacket files.

Actifilm (CIM-Card). Actifilm is a special Mylar or acetate film sheet that has been dye-impregnated with a Diazo compound, which is sensitive to the ultra-violet spectrum and can be developed with ammonia vapor.[11] The original system design and exploitation of this process was carried out by Douglas Aircraft Company, who coined the name Continuous-Image Microfilm Card (CIM-Card) to describe the basic working medium. The film stock and equipment were developed by the Ozalid Division of General Aniline and Film Corporation. The primary advantage of Actifilm is that the entire area of the card, which comes in many sizes, can be used for the storage of an image; the image is a part of the card stock, and will withstand extensive use without scratching.

[11] Jones, R. E., "Actifilm in Action," *American Documentation,* Vol. 12, No. 3, pp. 222–223 (July 1961).

Nivens, F. A., "Actifilm and the CIM Card," paper presented at the Los Angeles Chapter of the American Documentation Institute Symposium on Information Processing, Los Angeles, California (September 1961).

Nivens, F. A., "CIM Card Cuts Service Office Piling Jam," *Industrial Photography,* March 1959, p. 104.

The aircraft industry is representative of the industries that have many drawings too large to be conveniently contained in one frame of microfilm, or mounted on one aperture card. Douglas Aircraft Company was frequently faced with the problem of microfilming original aircraft loft drawings from 10 feet long to, sometimes, 100 feet long. It is more convenient and more efficient if the entire drawing is seen as a continuous, rather than a segmented, image.

In addition to providing long continuous images such as large engineering drawings (see Fig. 9-13), several frame sequences, or unusually long frames of microfilm can be transferred to a single film card instead of being segmented on several different cards. Figure 9-14 shows a single 5-by-8-inch Actifilm sheet with 224 images of the California Appellate Reports transferred from 16mm roll film.

Actifilm stock is also available in the same dimensions as punched tabulating cards for IBM or Remington Rand machines, and can be used in the viewing and copying equipment in the same manner as the aperture card. The basic Actifilm card stock has a higher storage capacity than the aperture card. For example, a 3-by-5-inch sheet could easily contain 30 documents, and a sheet of punched card dimensions could easily contain 16 to 70 page images, depending on the reduction ratio. In actual practice, it may prove advantageous to use a portion of the card to hold the indexing information or auxiliary information. The ability to arrange the images (and pertinent information) in several ways on the card stock offers a great deal of flexibility. For example, a convenient arrangement in practice is to put an engineering drawing on one part of the card, and all of the pertinent engineering change orders or revision notices on the other. For the card user, this provides a very concise grouping of the drawing with its related change orders or other documents.

At present, one of the main drawbacks to the use of Actifilm is the comparatively high cost of the Actifilm stock. However, there may be many applications in which the special features and characteristics of this system outweigh the cost disadvantage.

***Microfilm** Sheets.* One method of providing unit-record microfilm copies is to place one or more microfilm images on transparent film stock of the same dimensions and stiffness as filing card stock. Such microcopies can be filed away as a unit-record in much the same manner as a catalog card. The card could contain provisions for manual or machine searching, and could probably be used with standard viewing or reproducing equipment.

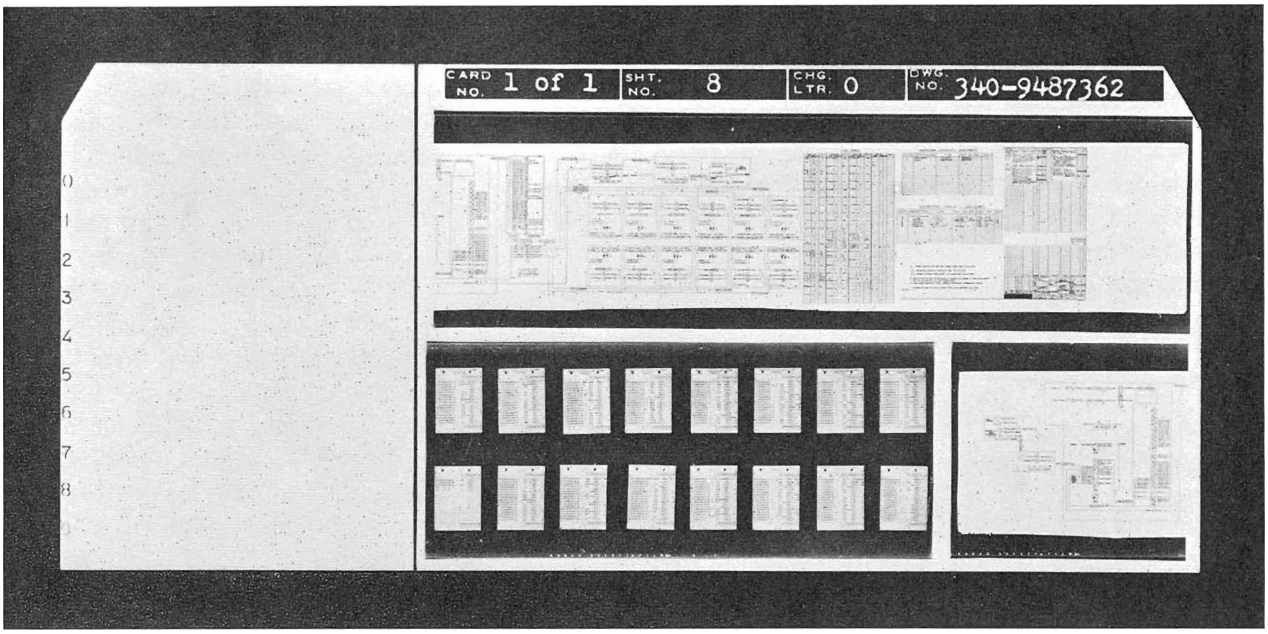


FIG. 9-13 Sample Actifilm sheet in tab card size (slightly smaller than actual size).

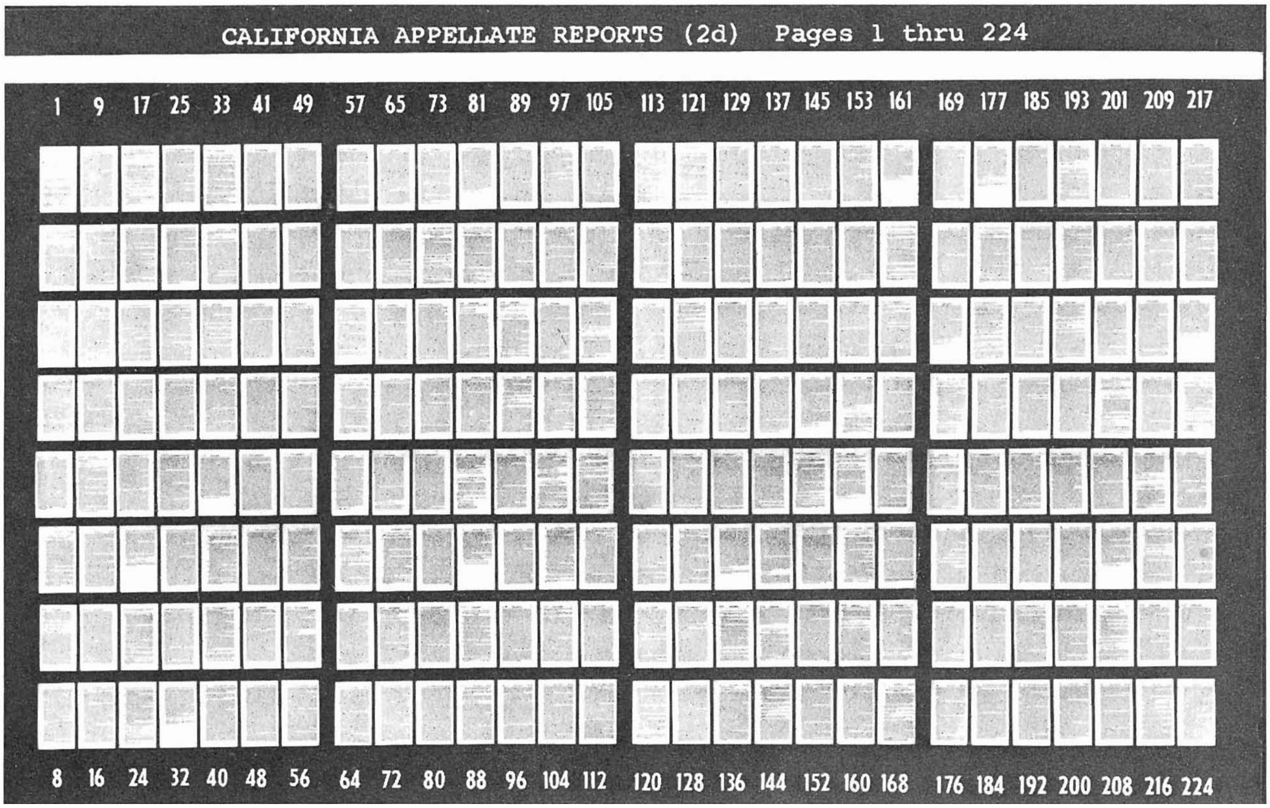


FIG. 9-14 Sample 5-by-8-inch Actifilm card.

One example of this type of card is that developed by the Petroleum Research Corporation (shown in Fig. 5-38) which puts an 18:1 reduction of a journal article onto a 5-by-8-inch card. These particular cards are furnished by Petroleum Research Corporation as part of a subscription library service for people who are interested in automatically receiving file copies of significant articles in particular subject fields. In this case, part of the card was arranged for interior needle sorting, but there was room enough on the remainder of the card to hold up to 56 pages of text. For convenience, the title of the article on the card is made large enough to read without viewing equipment.

Opaque Microcopies. Two main forms of opaque microcopy are Microcards and Microtape.[12] In both cases, the storage medium is a positive photographic print, made from film negatives, that will remain clearly legible for archival purposes, and will neither smudge nor rub off. The master film negative can be used to generate as many copies as required. The storage medium is opaque, and must be viewed by reflected instead of transmitted light.

Microcards are currently available as 3-by-5-inch cards, which can easily hold 70 page images on each side (see Fig. 9-15). The reduction of the original image on the card is usually between 17:1 and 23:1. Cards with images on one side only can be shelved at about 85 cards to the inch, and cards with images on both sides (containing up to 140 pages) can be shelved at about 65 to the inch. They can be stored in regular card files, and there are no special environmental restrictions.

Microtape consists of rolls of 16mm or 35mm positive microcopy with a pressure-sensitive adhesive coating on the back (see Fig. 9-16). Individual frames may be cut at desired intervals from this roll and pasted onto other papers or records in much the same manner as adhesive labels. A 3-by-5 filing card for example, can easily hold 40 page images. Viewing equipment is available

[12] Haas, W. de, "The Microfiche," American Documentation, Vol. 9, No. 2, pp. 99–106 (April 1958).

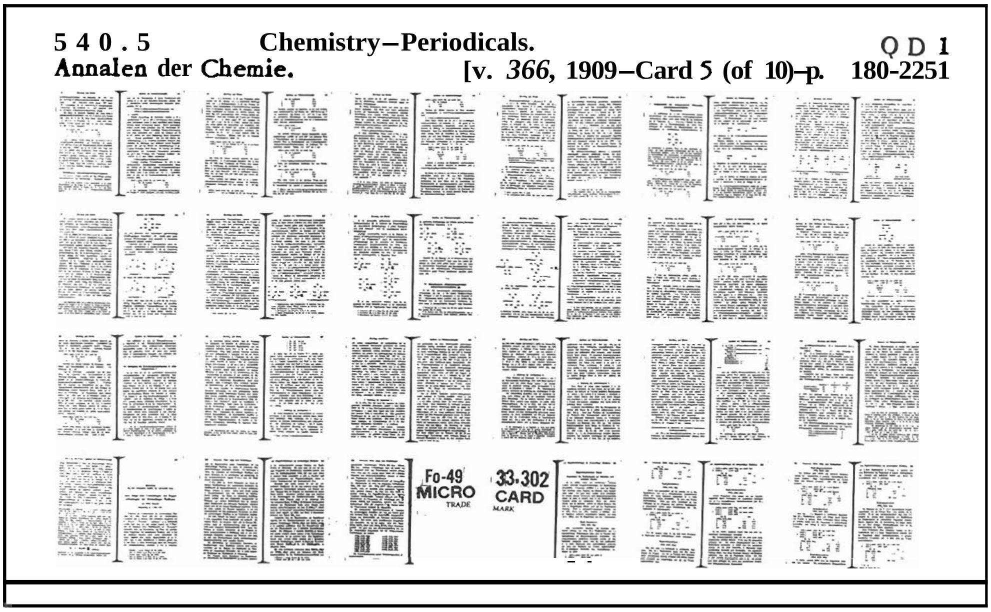


FIG. 9-15 Representative Microcard.

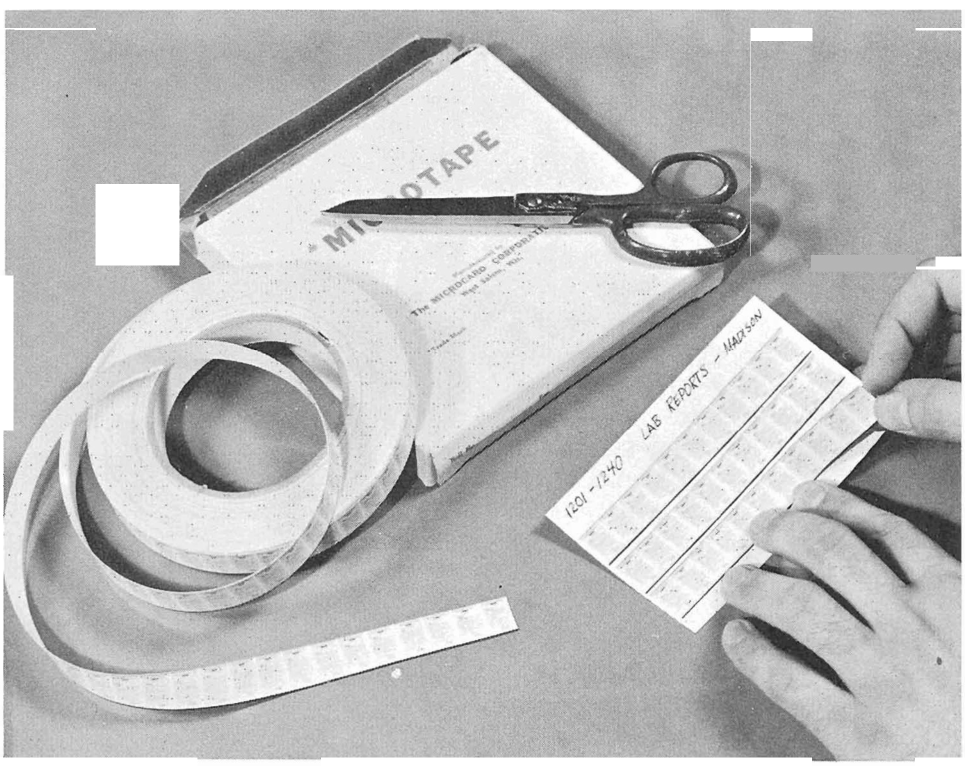


FIG. 9-16 Microtape.

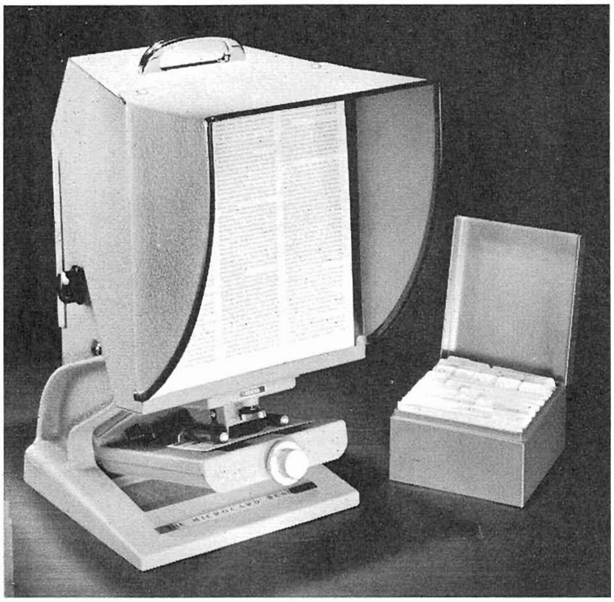


FIG. 9-17 Microcard reader (Model Mark VII).

(see Fig. 9-17) as well as equipment to produce page-size hard copy from Microcards or Microtape (see Fig. 9-18). The cost of a 100 foot roll of this tape (excluding the cost of the original microfilming) is about $9.50.

MECHANIZED MICRO-IMAGE SYSTEMS

A good share of the equipment development effort for information retrieval problems has been directed toward the design and development of complete systems for the storage and retrieval of information in micro-image form, whether copies of photographs, drawings, or documents. However, only a small amount of equipment is in operation—primarily for use in government installations.

In all of these systems, the file item usually takes one of two general forms: (1) an image with associated indexing information used for machine searching (in unit-record or continuous-record form); and (2) an image at a specific address or location without any indexing information physically associated with the image (indexing information exists, but in a manual or mechanized system separate from the image storage). The unit-records can be mechanically and logically independent, with both indexing and graphic information stored on each film chip or card. The unit-records can also be logically independent, but mechanically constrained in serial order with both indexing and graphic information stored on each film frame, such as with indexed roll microfilm.

Most of the mechanized image systems operate with reductions of approximately 30:1, although some use much greater reductions (e.g., the Minicard equipment with 60:1, the Verac equipment with 70:1 or 140:1, and the photochromic micro-image equipment with 200:1), and the technology is available to reduce images to a much greater degree, say 1000:1.[13] In nearly all cases, the image is stored on some type of film emulsion (e.g., silver, Kalfax, Diazo). In a few cases, the image is stored on magnetic tape by television recording techniques.

[13] Heilprin, L. B., "Communication Engineering Approach to Microforms," *American Documentation,* Vol. *12,* No. 3, pp. 213–218 (July 1961) and *Proceedings of the Tenth Annual Convention of the National Microfilm Association,* pp. *80–92* (April 1961).

FIG. 9-18 Microcard copier (Model 1).

Images Accompanied by Indexing Information

Microfilm Roll Systems

THE RAPID SELECTOR. The Rapid Selector was one of the earliest microfilm viewer-searchers.[14] It was developed in 1949 by Engineering Research Associates under the sponsorship of the Office of Technical Services of the U.S. Department of Commerce. The Rapid Selector used 2000-foot rolls of microfilm with the indexing data close to the film image. Each film frame normally stored the reduced image of a single page. The system had the capability of storing 72,000 frames on a single reel, with 216 positions of binary coding for each frame. Searching was done at the rate of 500 feet per minute (330 frames per second). Frames that satisfied the search criteria were copied on another roll of microfilm by a high-speed intermittent camera, so that the immediate result of the search was a roll of exposed microfilm. No convenient facilities were available to project or view the selected frames, making it impractical to browse in the files or to modify a search already under way. Several coding systems were proposed for this device by Wise, Perry, and Mooers, and their comments[15] are generally applicable to current or proposed viewer-searchers.

Yale University later redesigned this machine to meet their special requirements for document searching. Shortly thereafter, in 1959, the U.S. Patent Office and the U.S. Navy Bureau of Ships asked the National Bureau of Standards to study the original machine to see if it could be modified and updated to meet their needs. As a result of this study, a new Rapid Selector was developed by the Bureau of Standards under Bureau of Ships sponsorship.[16] The new unit (see Fig. 9-19) searches the 35mm film at a rate of 6000 pages per minute. There are 240 binary indexing positions for each frame (see space for 240 dots between frames in Fig. 9-20), with 40,000 pages per reel, at an 8:1 reduction. Binary notation and superimposed coding are used extensively. The coding information includes the usual bibliographic information, and major and secondary subject-matter designations. Selected frames are photographed on a strip of recopy film, and a number of consecutive frames can be copied without slowing or stopping the master

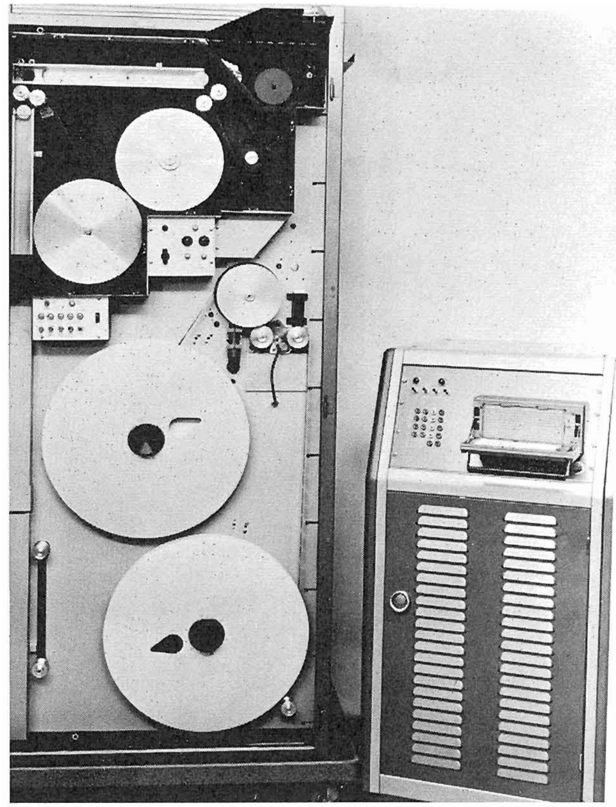


FIG. *9-19* Bureau of Ships Rapid Selector.

[14] Shaw, R., "The Rapid Selector," *Journal of Documentation,* Vol. 5, No. 3, pp. 164–171 (December 1949).

Shaw, R., "High Speed Intermittent Camera," *American Documentation,* Vol. 1, No. 4, pp. 194–196 (October 1950).

Green, J., "The Rapid Selector, An Automatic Library," *Reviews of Documentation,* Vol. 17, p. 66 (1950). (Reprinted from the September 1949 *Military Engineer.)*

Anon., "Photoelectric Librarian," *Electronics,* Vol. 22, No. 9, p. 122 (September 1949).

[15] Mooers, C., "Coding, Information Retrieval, and the Rapid Selector," *American Documentation,* Vol. 1, No. 4, pp. 225–229 (October 1950).

Wise, C., and J. Perry, "Multiple Coding and the Rapid Selector," *American Documentation,* Vol. 1, No. 2, pp. 76–83 (April 1950).

Wise, C., and J. Perry, "Multiple Word Coding vs. Random Coding for the Rapid Selector," *American Documentation,* Vol. 3, No. 4, pp. 223–225 (October 1952).

[16] McMurray, J. P., "The Bureau of Ships Rapid Selector System," *American Documentation,* Vol. 13, No. 1, pp. 66–68 (January 1962).

Bagg, T., and J. Pike, "The Rapid Sklector and Other National Bureau of Standards Document Retrieval Systems," *Proceedings of the 11th Annual Convention of the National Microfilm Association,* pp. 213–228 (National Microfilm Association, Annapolis, Maryland, 1962).

National Bureau of Standards, Washington, D.C., *The Rapid Selector—An Automatic Document Retrieval Device,* Summary Technical Report, STR-2388 (September 1959).

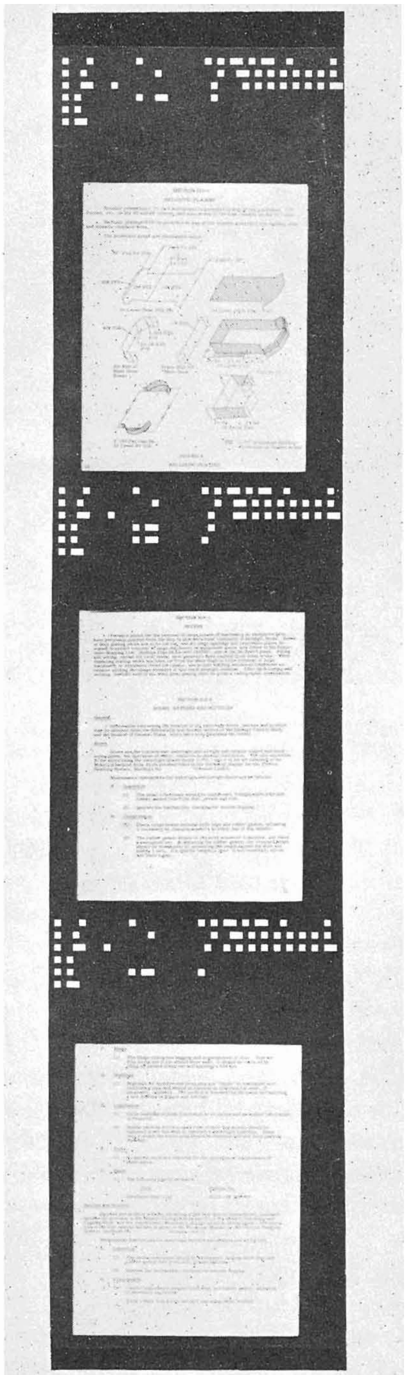


FIG. 9-20 Sample film and coding used with the Bureau of Ships Rapid Selector.

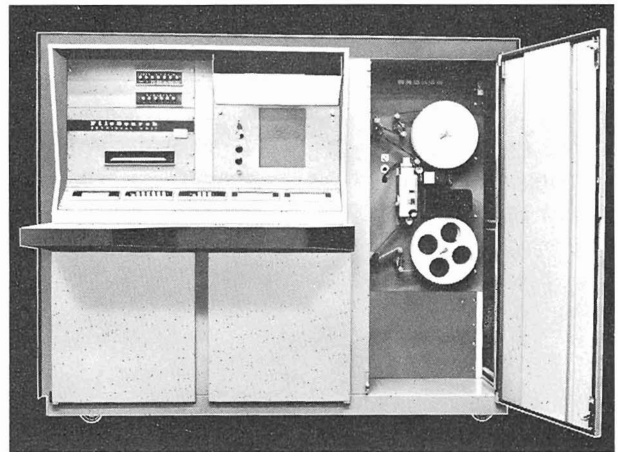


FIG. 9-21 FMA Filesearch retrieval unit.

film. A predictive search, in which copies are not made, is used to evaluate the question. The immediate result of a search is the re-copy film, as in the original Rapid Selector. The average time for a complete search is 12 minutes, and the present system is currently performing an average of 150 searches per month.

FMA FILESEARCH SYSTEM. This equipment (see Fig. 9-21) is an up-to-date commercial version of the Rapid Selector, and has some interesting system features. The file medium is a 1000-foot roll of 35mm film, with a maximum of 32,000 document pages per reel. The pages are recorded at a reduction of 25:1, and optically encoded indexing information accompanies each film frame. With this reduction ratio, 32 legal-size pages can be recorded on each foot of film. It is estimated that a single camera operator can photograph about 4000 pages a day.[17] Space is provided on the film for a total of 56 alphabetic or 84 decimal characters of indexing data on each frame, plus some marks for machine checking (see Fig. 9-22). The indexing data can take any general alphabetic or numerical form.

The system consists primarily of a recording unit and a search unit. The recording unit uses a 35mm planetary camera to record both the document page and the indexing data on the same frame at the same time. To speed up the generation of the indexing code pattern for filming, the indexing data

[17] Condon, R. A., "The FMA Filesearch System," paper presented at a Symposium on Information Processing, Los Angeles Chapter of the American Documentation Institute, Los Angeles, California (September 1961).

are first put in punched card form, so the cards can subsequently be used to set up the patterns quickly and automatically during the filming operation.

The search operation consists of typing a search descriptor or index number in a punched card, then using the card to set up the search pattern in the machine. The system is capable of simultaneous handling of up to 6 requests having a moderate degree of logic complexity. The file is searched at the rate of 6400 pages (200 feet) per minute, and at the user's option, the selected frames are either displayed on a viewing screen, printed as full-size hard copies, or copied on another reel of microfilm. Five minutes or more would be required to search the entire 32,000-page file. One of these units is currently in operation with Central Records Section of the Bureau of Ships.

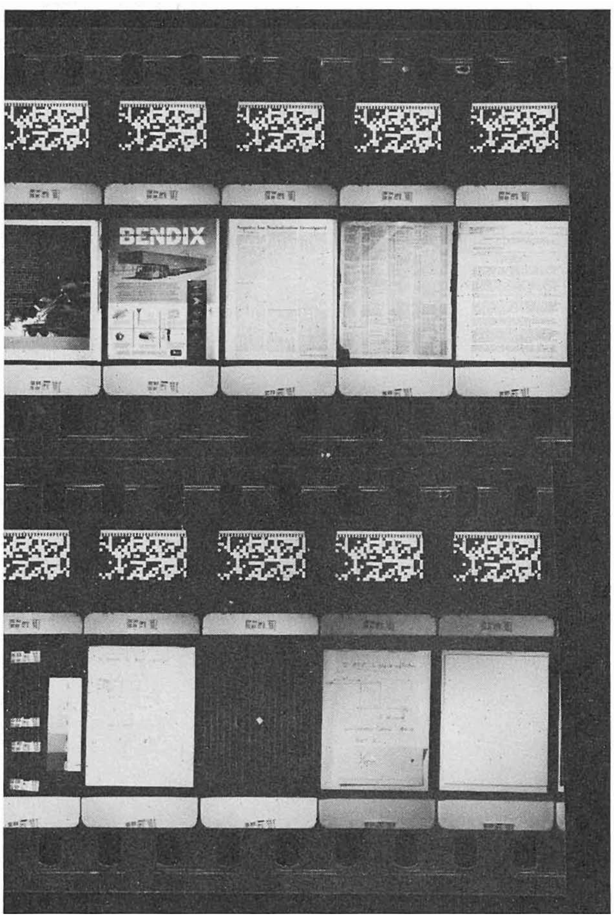


FIG. 9-22 Sample film and coding used with the FMA Filesearch retrieval unit.

FIG. 9-23 Benson-Lehner FLIP.

BENSON-LEHNER FLIP. The Benson-Lehner FLIP (Film Library Instantaneous Projection) equipment currently operates as an automatic selector for viewing purposes (see Fig. 9-23).[18] Each frame of 16mm microfilm contains a single image and 32 bits of binary coded indexing information to describe the contents of that frame. A 1200-foot roll with 72,000 images is searched at a rate of 300 to 600 frames per second. The minimum scan speed is 60 inches per second. The equipment for the operator consists essentially of a keyboard interrogation device, and a large display screen. The keyboard positions correspond to the coding positions on the film so that the inquiry can be keyed into the system via the keyboard. After the inquiry has been entered, the film is transported at a high rate of speed until a frame is located which satisfies the search criteria. The film is then stopped and positioned so that the chosen frame is displayed on the viewing

[18] Worsley, P. K., "Data Retrieval with Especial Application to Use of Film Library Instantaneous Presentation (FLIP) in Literature Searching," a section of the book *Modern Trends in Documentation*, M. Boas, editor, pp. 70–73 (Pergamon Press, New York, 1959).

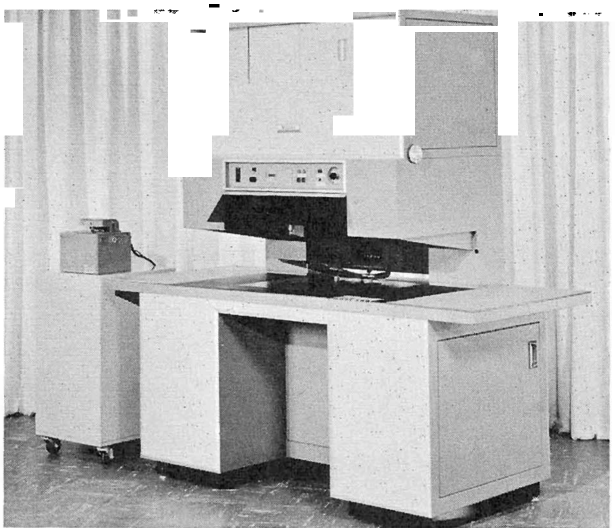


FIG. 9-24 Minicard document camera.

screen. No provisions have been made for reproducing the selected frame, and no general-purpose equipment is currently available to prepare the original film record with the coded indexing information.

Microfilm Unit-Record *Systems*

THE EASTMAN KODAK MINICARD SYSTEM. The Minicard system (see Figs. 9-24 through 9-28) was developed by Eastman Kodak with U.S. Air Force support for application to specific military information problems rather than for application to documentation problems in general.[19] It comprises one of the largest and most expensive collections of equipment ever developed specifically for information storage and retrieval. Primarily because of its cost, its use has been restricted to a few special government file problems. The first complete system went into operation late in 1958.

In this system, the unit-record or film chip is a 16mm-by-32mm piece of film which normally includes one or more images and their associated indexing data (see Fig. 9-29). Images are stored at a reduction of up to 60:1 and up to 12 legal-size pages can be recorded on a single chip. Single maps up to 18 by 22 inches can be recorded at a 38:1 reduction, and single photographs up to 9 by 9 inches can be recorded at a 20:1 reduction. The amount of indexing data on each chip may range from 252 to 2730 bits, roughly corresponding to 42 to 455 alphanumeric characters. Because of their small size, it has been found convenient to carry the chips on a skewer in groups of 2000 or less when manual handling is necessary.

The system includes a camera that photographs the original documents on 16mm film and inserts the coding information on the same frame as the related image, using data provided on punched paper tape. Approximately 500 pages of graphic material can be recorded in an hour, assuming an average of 6 pages and 210 characters of code on each film chip. A 200-foot-film magazine will record approximately 1800 film records. After the film is processed (about 30 minutes per magazine) and checked for image quality, it is chopped into individual chips by a film cutter that can cut an entire film magazine into about 2000 chips in less than 4 minutes. These original film negatives are used to produce second-generation positives for the working file. The original negatives are then put away for security storage and are not used for any file searching operations. The second-generation positive chips are produced in an automatic duplicator, one for each input document, and one for every subject heading or file category to be used. They are then inserted into the file under the respective categories. This procedure serves to partition the file so that searches can be confined to scanning selected portions of the file, instead of requiring a scan of the complete contents. The input chips are automatically routed to their respective magazines at the rate of 1000 chips per minute by a sorting machine.

[19] Tyler, A. W., W. L. Myers, and J. W. Kuipers, "The Application of the Kodak Minicard System to Problems of Documentation," American Documentation, Vol. 6, No. 1, pp. 18–30 (January 1955).

Kuipers, J. W., A. W. Tyler, and W. L. Myers, "A Minicard System for Documentary Information," American Documentation, Vol. 8, No. 4, pp. 246–268 (October 1957). This is also published as Chapt. 27 in Information Systems in Documentation, J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

Anon., "More Instant Literature" (Minicard description), Chemical and Engineering News, June 29, 1959, pp. 82–83.

Myers, W. L., and G. L. Loomis, "The Minicard Film Record as a Common-Language Medium," Chapt. 20 in *Information Retrieval and Machine Translation*, Part 1, A. Kent, editor (Interscience Publishers, New York, 1960).

Effros, A. L., "The Minicard System for Storage and Retrieval of Documentary Information," paper presented at a Symposium on Information Processing, Los Angeles Chapter of the American Documentation Institute, Los Angeles, California (September 1961).

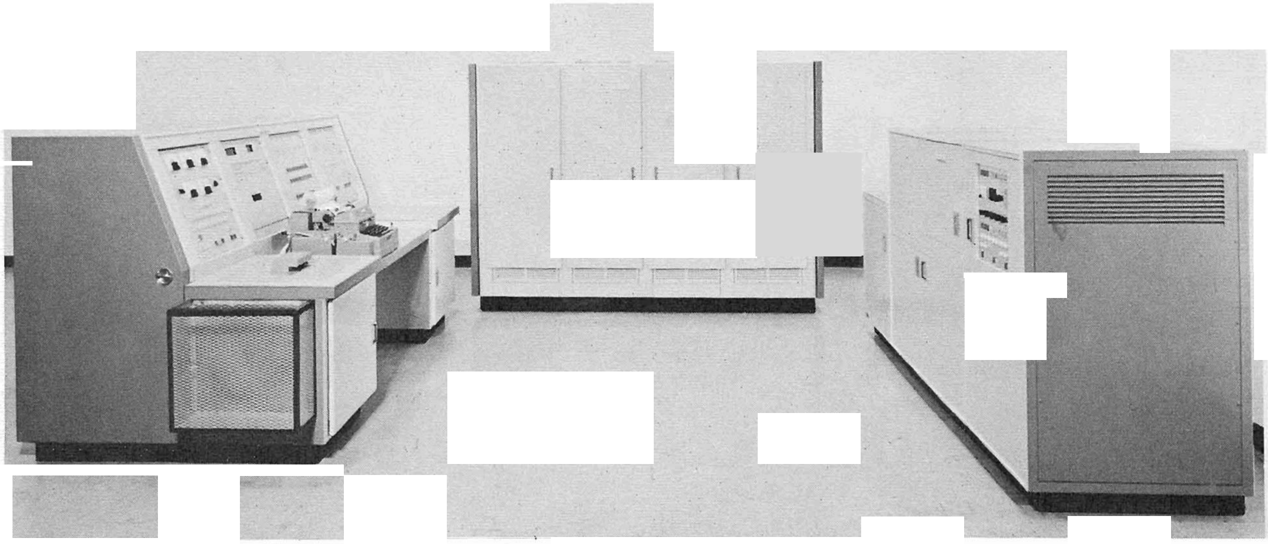


FIG. 9-25 Minicard computerduplicator.

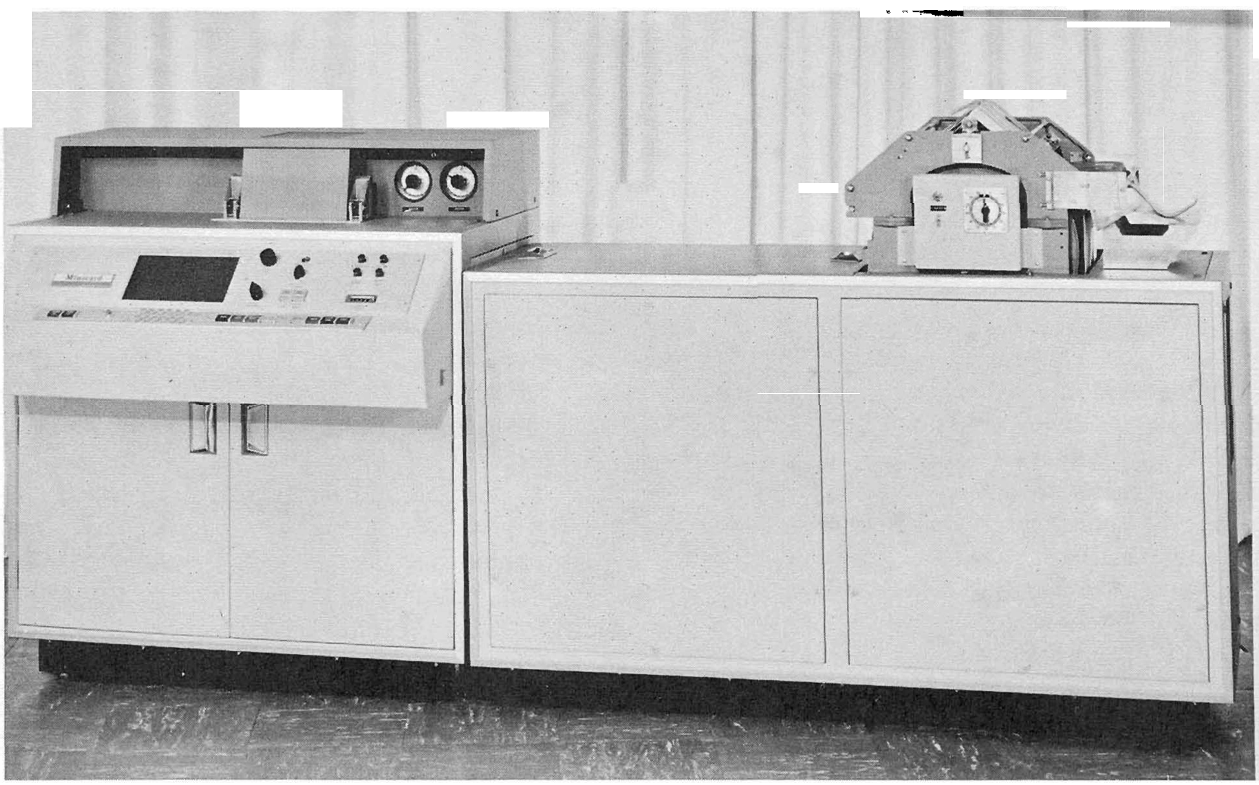


FIG. 9-26 Minicard document enlarger.

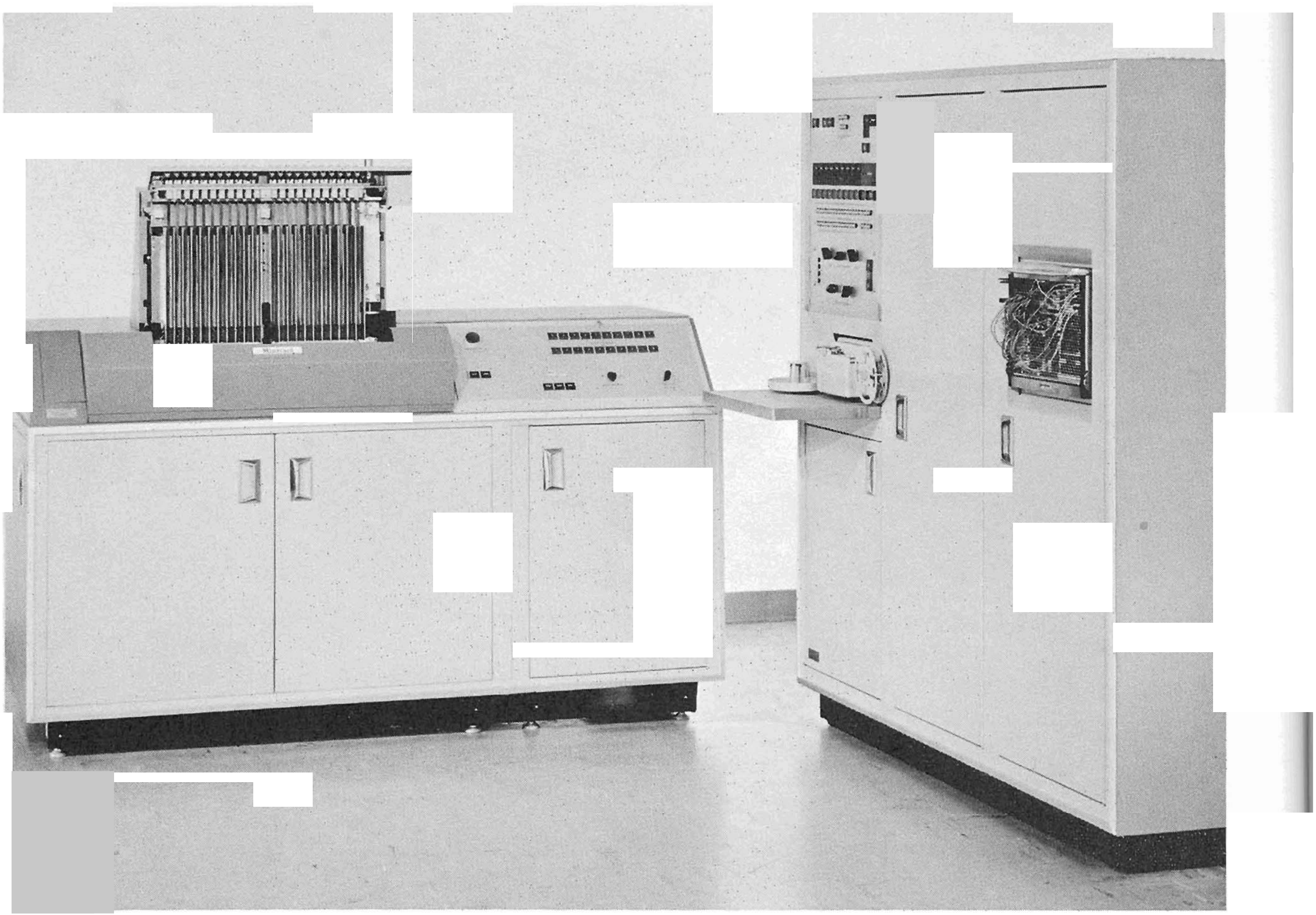


FIG. 9-27 Minicard selector-sorter.

File searching is done by an automatic selector at the rate of 1200 chips per minute. The question statements are supplied on paper tape and the question logic is stated with plugboard wiring. The records that satisfy the search criteria are copied on third-generation negatives and delivered to the inquirer. Equipment is available for viewing, or for producing hard copy at the rate of 540 prints per hour. The film chips cannot be conveniently handled, viewed, or reproduced by any equipment except the equipment developed for the Minicard system.

THE FILMOREX SYSTEM. The Filmorex system is a photographic unit-record system which has been under development since 1952 by Dr. Samain in Paris, France.[20] Over 20 of these systems have been delivered to European users. One representative operating installation is shown in Fig. 9-30. The unit-record (see Fig. 9-31) is a film chip (60mm lengths of 35mm film) with half binary coding and half text, similar to the format used in the Minicard system. Special equipment is available to search, view, or produce hard copy photolistings of the cards. Documents up to 30 by 45 centimeters (approximately 12 by 18 inches) can be photographed on one half of the microfilm card. The coding half of the card will hold a binary pattern for up to 25 sjx-digit numbers that can be used to index or describe the contents of that card.

The original microfilming is done by photographing "coordinate cards" with a special camera unit

[20] Samain, J., "Documentation by the Filmorex Technique," Chapt. 26 in *Information Systems in Documentation*, J. H. Shera et al., editors (Interscience Publishers, New York, 1957).

FIG. 9-28 **Operating Minicard installation.**

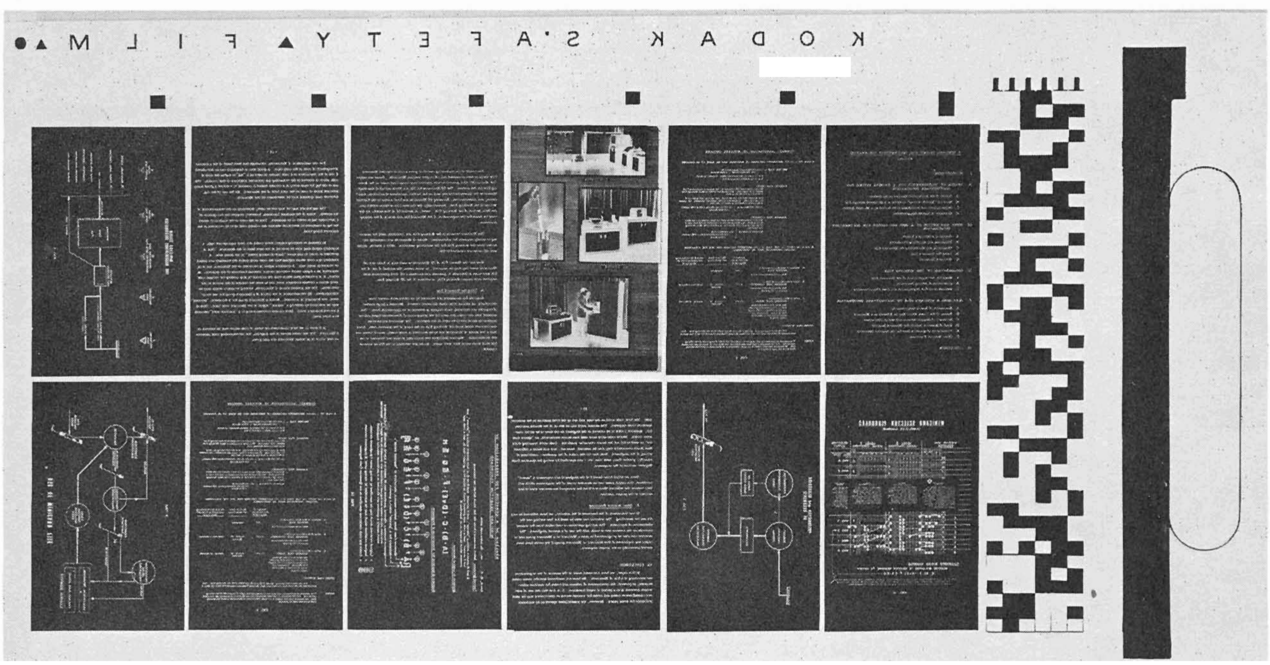


FIG. 9-29 **Representative Minicard film record (enlarged).**

(see Fig. 9-32). The coordinate cards are cards for each indexing term which have the appropriately coded pattern printed on the edge. Because of this, the digital coding for a document could be displayed for photographing by shingling the appropriate coordinate cards next to the document. This is a very simple method for inserting the coding terms on a document, and could certainly be applied to other microfilm systems. The searching is performed by a selector unit which reads the cards at a rate of 600 cards per minute (see Fig. 9-33). When the selector recognizes a coding combination that it has been instructed to watch for, the card containing this code is shunted into a separate pocket. A hard copy listing of the selected cards may be obtained by running the cards through the photolister unit (see Fig. 9-34).

Video Storage. An image storage system designed by RCA, and referred to as the Video File stores graphic images such as documents, maps, or drawings on video magnetic tape in much the same manner as a video tape recording system. Each frame or image on the tape is accompanied by digital data giving the file number or address of the image. With a resolution four times greater than commercial television, this system can store a page-size document on about 3 square inches of video tape surface. A 7200-foot reel of tape could store up to 36,000 pages. The documents are initially recorded on tape by a scanning TV camera which, with automatic page-feeding equipment, can operate at about 120 documents per minute. The file tape can be searched at the rate of 300 inches per second, and selected images can be displayed on a monitoring screen or printed out by an Electrofax printer. This equipment has been proposed but not used yet for any operational information system.

Images with Addresses

Several pieces of equipment have been developed to store images and recall them, given specific addresses or file numbers. The search for appropriate document numbers is done beforehand by any sort of manual or machine method and the only job for the storage machine is to provide a display or copy of the document requested by number. Such a machine works with any type of indexing or classification system, and any type of search mechanism can be used to obtain the relevant document numbers.

Recordak Lodestar with Counting Accessory. The regular Recordalc microfilm viewer can be augmented with a special accessory that counts the

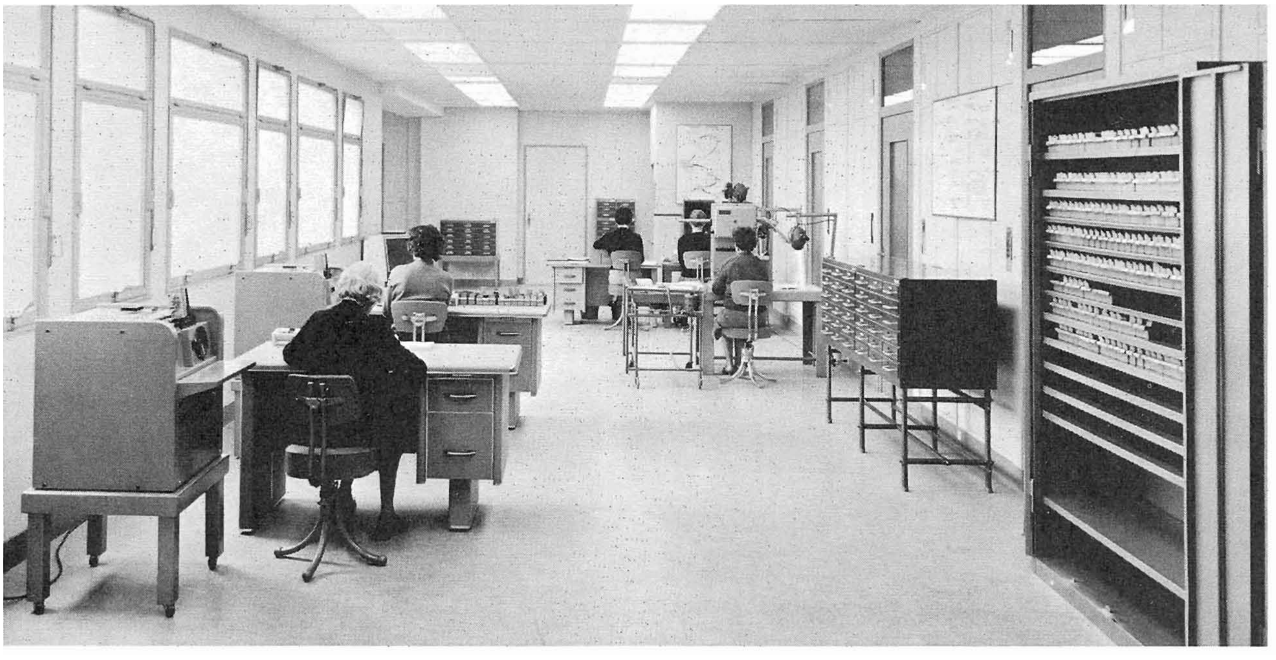


FIG. 9-30 Operating Filmorex installation.

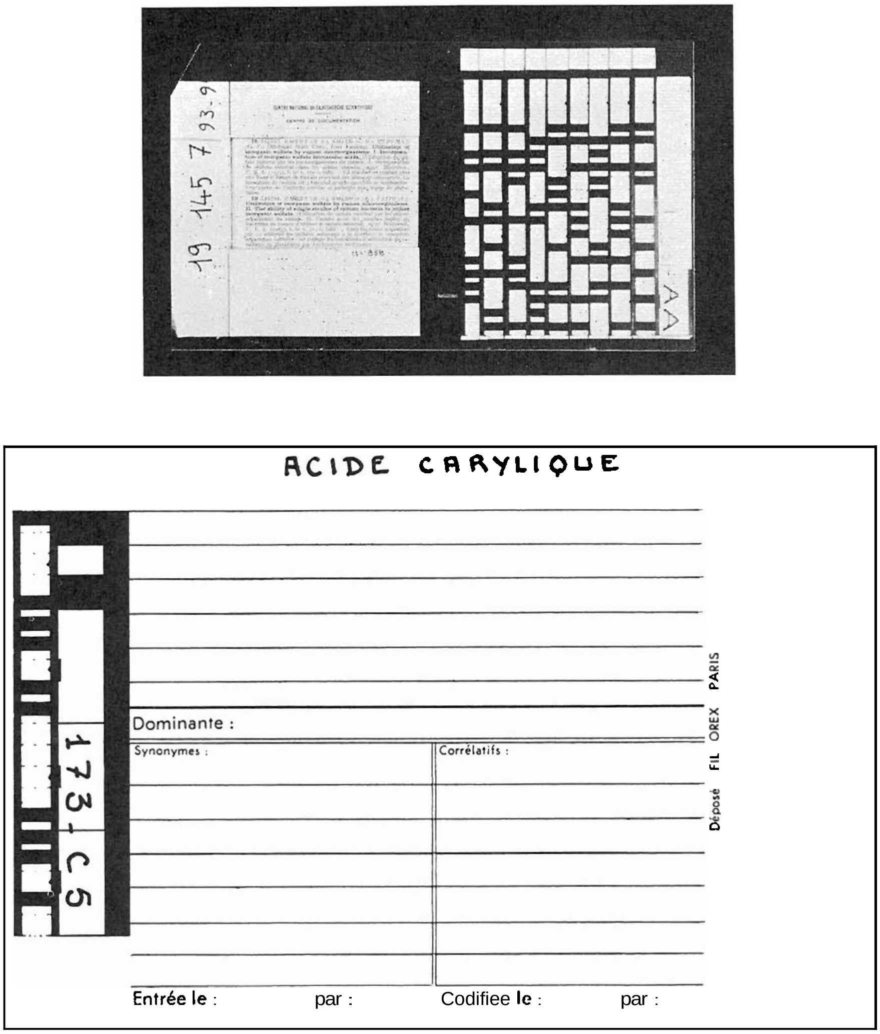


FIG. 9-31 Representative Filmorex film record and coordinate card.

number of frames since the beginning of the reel. If the location or sequence number of a document on a roll of film is known, the number can be entered on a keyboard, and the drive motor set to unwind the film automatically until it arrives at or near the desired address. The operator need not monitor the viewing screen during this scanning operation, so is free to concentrate on other tasks. This file look-up device is admittedly crude, but it is relatively inexpensive.

Ferranti-Paclcard Rapid Access Look-Up System. The Ferranti-Packard equipment (see Fig. 9-35) scans a continuous loop of 16mm film at moderate speed (over 40 inches per second) to achieve a moderate random access time. One installation has a single loop of 440 film frames, giving an average frame selection time of 1.5 seconds. Larger loops are possible with consequent increase in time; for example, a 5000-frame loop would give an average access time of about 18 seconds. Each film frame contains space for a small amount of indexing data, arranged in much the same manner as in the Rapid Selector. The search criterion is a single number, such as a page number, and is entered on a keyboard device. The selected frame is displayed on a viewing screen. Equipment is being developed to print full-size copy from the selected frames.

One of the earliest uses of this equipment was by a wholesale drug company's sales clerks who received and wrote up telephone orders. A stock number and unit price had to be looked up for each

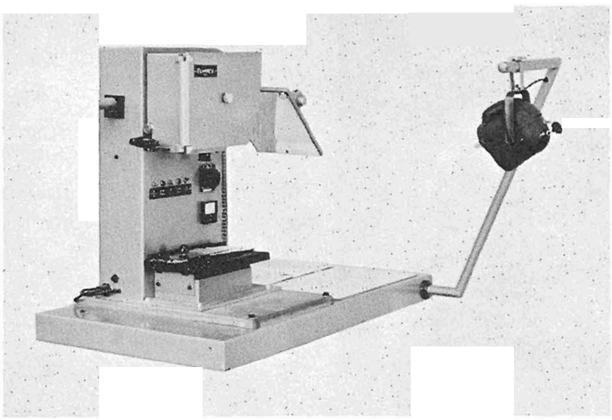


FIG. *9-32* **Filmorex** camera.

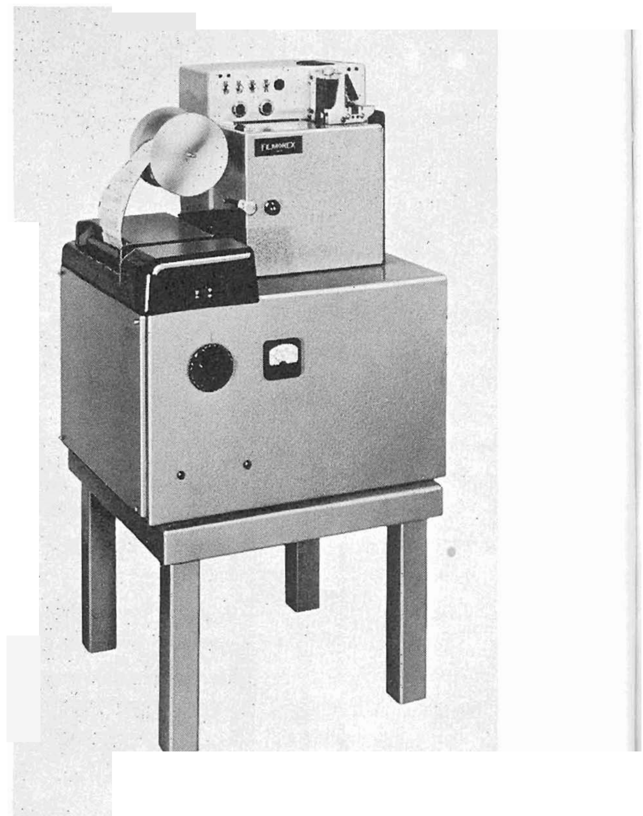


FIG. *9-34* **Filmorex** photolister.

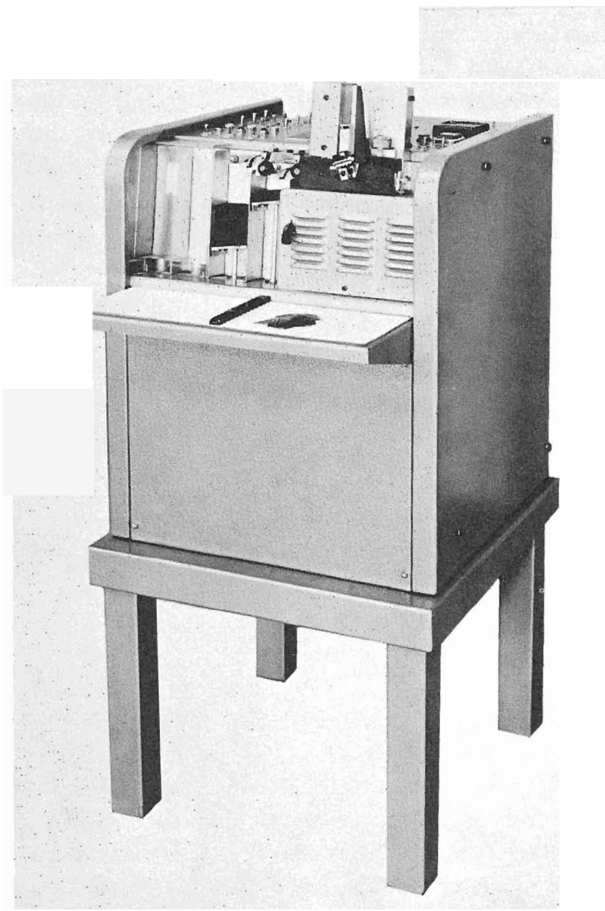


FIG. *9-33* **Filmorex** selector.

item ordered, which had previously been done with a conventional tub file. With the microfilm equipment the clerks needed only to enter a number in a keyboard to cause the stock number, unit price, and other data to be displayed on a screen, much faster than they could be obtained from the tub file.

CRIS. The CRIS (Command Retrieval Information System), developed in 1962 by a subsidiary of Information for Industry, is an outgrowth of some earlier work done by Avakian in 1956.[21] This system stores images photographically on a scroll of microfilm 400 feet long by 17 inches wide; each scroll contains over 500,000 page-size images or over 28,000 large drawings. **A** keyboard device is

[21] Avakian, E., and E. Garfield, "AMFIS—The Automatic Microfilm Information System," *Special Libraries,* Vol. *48,* No. 4, pp. 145–148, April 1957.

Larson, P. W., "CRIS (Command Retrieval Information System)," *Proceedings of the 11th Annual Convention of the National Microfilm Association,* pp. *41-50* (National Microfilm Association, Annapolis, Maryland, 1962).

FIG. 935 Ferranti-Packard rapid access look-up system.

used to enter a CRIS address, and the image at that address is displayed to the operator or is provided in aperture card form (see Fig. 9-36). The average time to retrieve any desired image is under 20 seconds, and the retrieval time for sequential images is even faster. An aperture card copy of the displayed frame can be made in about 20 seconds. The scroll material is a Mylar base with a Kalfax emulsion and is usually prepared by contact printing from strips of microfilm that were obtained by conventional procedures.

Magnavox MEDIA. The MEDIA (Magnavox Electronic Data Image Apparatus) system was first demonstrated by Magnavox in 1961.[22] This photographic unit-record storage system uses a combination of manual and machine search techniques. The basic film chip is a 16mm-by-32mm card that contains up to two 9-by-15-inch page images or three 8%-by-11-inch page images at a reduction of 30:1, and an information field of up to 17 encoded and human-readable digits to identify the image. These digits usually give the document number as well as some other codes. The documents are photographed on 100-foot rolls of microfilm using a special camera unit (see Fig. 9-37), processed, and cut into cards at the rate of 240 per minute. The cards are kept in capsules or cartridges containing up to 200 cards. These capsules are stored in regular file cabinets and are handled and selected manually (see Fig. 9-38). The chips are usually not stored in order within a capsule but the contents of each capsule may be run through a sorting device (see Fig. 9-39) to extract specific chips. To retrieve a specific file item (e.g., document No. 1234598) a clerk selects capsule number 12345 from the file cabinet, mounts it in the selector unit, and keys in the last two digits (98). The cards are then scanned at the rate of 600 cards per minute to select the desired one. This unit can provide a page-size blowback of the card image if desired. A browsing unit is also available; with this, the user can view the cards in a capsule one at a time in a regular microfilm viewer-printer, and can transfer selected copies to hard copy form.

FIG. 9-36 CRIS (Command Retrieval Information System).

[22] Jenkins, D. D., "Magnetic Indexing, Microfilm Storage and Information Retrieval," in *Proceedings of the 11th Annual Convention* of *the National Microfilm Association,* pp. **205–212** (National Microfilm Association, Annapolis, Maryland, **1962).**
Laurent, R. L., "Magnacard—Magnavue—MEDIA," paper presented at a special conference of the Los Angeles Chapter of the American Documentation Institute (Los Angeles, September **1961).**

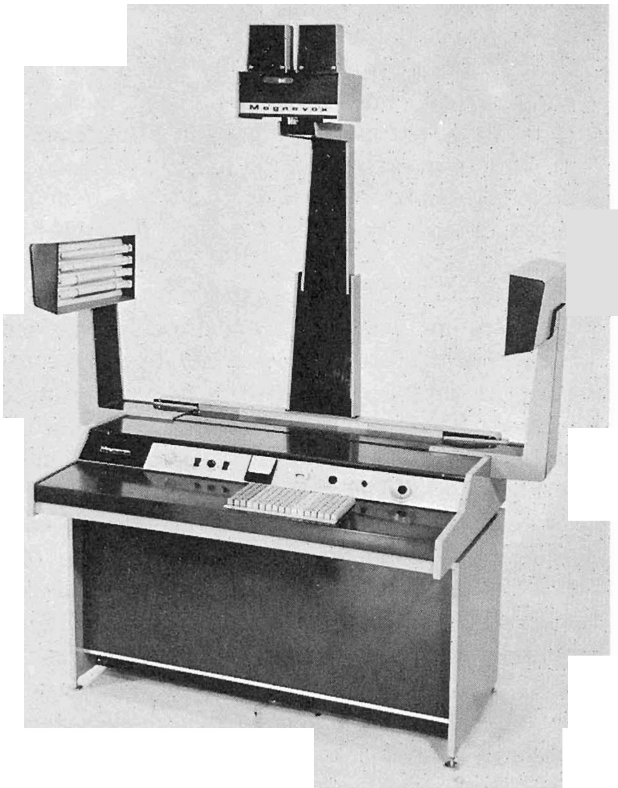


FIG. 9-37 Magnavox MEDIA camera coder.

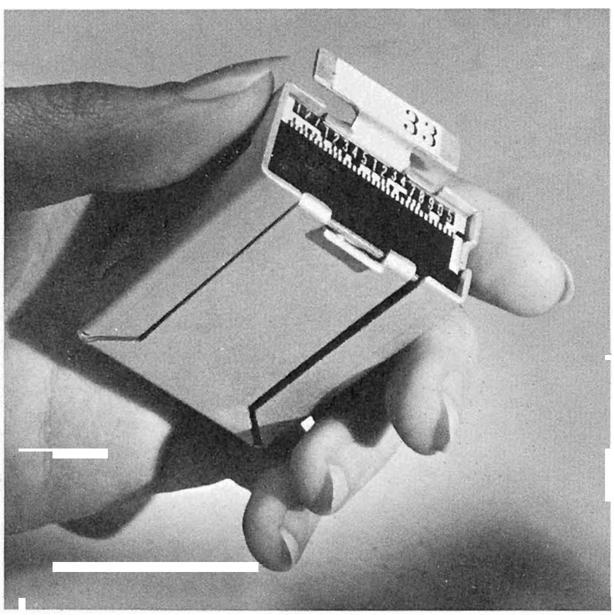


FIG. 9-38 Sample MEDIA capsule, showing a coded MEDIA card.

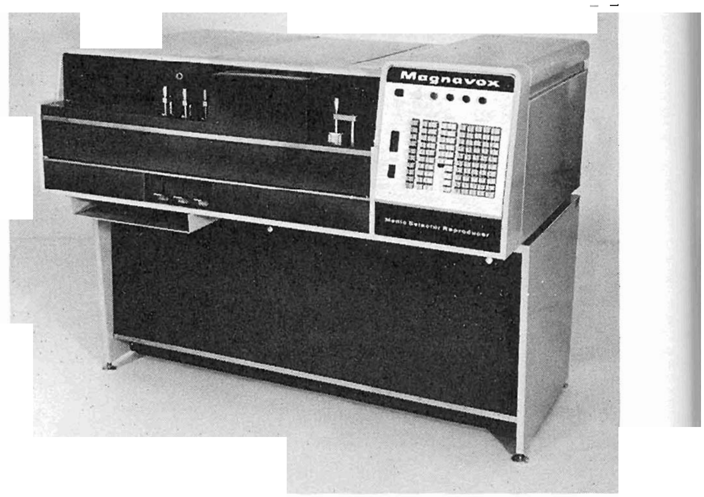


FIG. 9-39 Magnavox MEDIA selector-reproducer.

The Magnavox Magnacard, described earlier, in Chapt. 8, may also be used as an image storage system by utilizing one portion of the card to hold a microfilm image and another portion to hold the indexing data recorded in digital form on the magnetic tape part of the card.[23]

AVCO *Corporation* VERAC 903. The AVCO Manufacturing Corporation developed a single prototype model of a mechanized photographic storage system in 1959 with support from the Council on Library Resources.[24] This system, the VERAC 903, has not been operated yet in a real library environment. The equipment was designed to alleviate some of the storage problems of conventional library systems. Three main pieces of equipment have been developed: (1) a microphotographic memory, (2) a camera system to generate micro-images from the input documents, and (3) an output system to display and reproduce selected portions of the micro-image file. The photographic storage element was designed for a capacity of 1,000,000 reduced page images, with an access time between 0.3 and 2.0 seconds to any page in the memory. The images are stored at relatively high reduction ratios, 70:1

[23] Laurent, R. L., cited previously, fn. 22.
Gelb, J., *Utilization of Magnacard as a Display Device*, Report R357 of the Magnavox Research Laboratories, Torrance, California (February 1961); AD-251 271.

[24] Bowker, K., et al., *Technical Investigation of Elements of a Mechanized Library System*. Final Report EW-6680 of the AVCO Corp., Crosley Div., Electronics Research Laboratory, Boston, Massachusetts (January 1960).

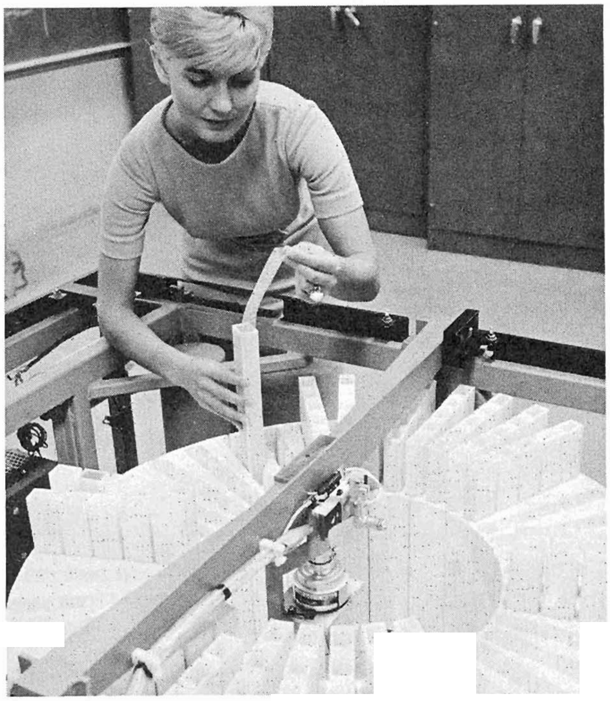


FIG. 9-40 WALNUT image file (IBM Model 9603 image file).

or 140:1. The output consists of a cathode ray tube display or a microfilm reproduction of the images selected from specific addresses or page numbers.

IBM WALNUT. WALNUT is a code name for a complex, mechanized, micro-image storage and retrieval system (IBM 9603 Image File) developed for a federal agency by the IBM Corporation.[25] The system consists primarily of a file of microfilm strips stored in bins and a mechanical selection device that can quickly go to a specified bin, mechanically select a strip of microfilm, and copy images from that strip onto aperture cards. The basic image file unit, shown in Fig. 9-40, contains a total of 990,000 page-size images (200 plastic cells of 50 strips each, with 99 images on each strip) stored at a 35:1 reduction. This is equivalent to about 3000 books or the contents of about 100 filing cabinets. Given an image number and page count, the random-access selection equipment can locate the desired image in less than 5 seconds and transfer it to a blank frame of Kalfax film mounted in an aperture card. The Kalfax film is exposed with an ultraviolet lamp and heat-developed by the basic file unit. The aperture card has positions for four images, reduced 27.2 times from the original document. The equipment can provide images from this file, on demand, at the rate of about 500 aperture cards per hour. The aperture cards can be viewed, or used to print copy on any of several models of commercially available microfilm viewer-printers. The original address of the image is given to the file in punched card form, and may be the direct result of a computer file search of some master index.

Another piece of equipment which accompanies the basic file unit is the IBM 9403 Image Converter (see Fig. 9-41). This unit serves as the file input device by transferring images from conventional 35mm sprocketed silver microfilm (at an original reduction of 15.8:1) to the Kalfax file strips at a further reduction of 2.2:1. The file strips utilize

FIG. 9-41 WALNUT film converter (IBM Model 9403 image converter).

[25] Porter, R. W., "A Large-Capacity Document Storage and Retrieval System," in *Large-Capacity Memory Techniques for Computing Systems,* M. C. Yovits, editor, pp. **351–360** (Macmillan Co., New York, **1962**).

Vogel, N. A., "WALNUT Document Storage and Retrieval System," *Proceedings of the 11th Annual Convention of the National Microfilm. Association,* pp. **27-39** (National Microfilm Association, Annapolis, Maryland, **1962**).

Bradshaw, P. D., "The WALNUT System: A Large Capacity Document Storage and Retrieval System," *American Documentation,* Vol. **13,** No. **3,** pp. **270–275** (July 1962).

Veyette, J. **H.,** Jr., "Photo-Image Storage: Its Role in Modern Business" (includes a description of the IBM WALNUT system), *Business Automation,* Vol. **6,** No. 4, pp. **1621** (October **1961**).

Kalfax film that has an image resolution on the order of 500 to 600 lines per mm. The Image Converter operates at a maximum rate of 1500 frames per hour, to transfer images to the file strips, assign addresses, and punch these addresses into a tab card.

A few of these **WALNUT** units have been developed, but to date only the Central Intelligence Agency is using them.

APPLICATIONS

General Considerations

System Studies. The microfilm equipment complement for any particular problem should be determined by a complete analysis of the system demands, and the economics of the proposed equipment. However, in many cases the economic factors will be obscured by the unavailability of accurate cost figures, or by the difficulty of pricing such an intangible factor as improved service. It may be difficult to justify microfilm systems solely on a basis of savings in personnel, material, or space rental costs. In all probability, one of the strongest justifications will be the improvement of existing services, or the ability to provide new services. On the other hand, the feasibility study that precedes the specification of equipment may reveal enough other possible changes and improvements in the present business system to make further mechanization unnecessary.

There are many microfilm systems in operation, however, and they are being used to solve a great variety of problems. In many cases, these systems have resulted in a savings in operating costs, or have improved the quality of the service. However, no equipment or systems are yet on the market that do an efficient job of filing and retrieving the film records. Major changes in components, storage and retrieval methods, systems design, material-handling techniques, and reproduction processes must be made before a collection of equipment can be proposed that is suitable for an integrated information-handling system. With the exception of the Minicard system, which is too expensive for all but a few applications, the most nearly mechanized system is the collection of equipment using the 35mm aperture card. However, it should also be noted that most organizations that use aperture cards do not exploit the characteristics of the tabulating card medium, and actually resort to manual file operations for fear of damaging the film image.

Although the applications discussed in this chapter do not represent integrated machine systems, it can be expected that within 5 or 10 years, such systems will become available, and will even become economically feasible for a great many applications. The development of microfilm systems, like that of information systems in general, will probably follow a cycle similar to the development cycle of electronic data processing equipment. Early data processing equipment was built to perform a very simple operation (compute), but was later developed to provide better communication facilities with the human users (high-speed printers, automatic-programing techniques, card readers, etc.), improved storage capacities (rapid-access storage and slower magnetic-tape or drum storage), and a capability for direct interrogation from remote locations (inquiry consoles). Further developmental effort and system studies provided means by which the information for the system could be generated at the source (e.g., punched paper tapes from cash registers) which could be used by equipment throughout all phases of the system. A great deal of effort was also expended in product planning, application studies, and investigations of better ways to use the equipment. A similar evolutionary cycle will probably occur for the microfilm systems. Early microfilm users only had a few pieces of single-purpose equipment (cameras, viewers, reproducers). Later systems were able to combine many of the functions into a single piece of equipment (viewer-printers) or to provide a certain degree of automatic operation (automatically printing the desired number of copies), or to use a common record medium for each of the equipments in the system (such as viewing, copying, or reproducing from aperture cards). Additional product planning, systems and component design, and applications studies will greatly improve the performance and flexibility of microfilm systems in the years to come.

Quality Control. Any comprehensive microfilm system must include some provision for checking the quality of the final film product. For engineering drawing systems it is usually necessary, prior to starting a microfilming program, to incorporate new drawing standards for such things as lettering and symbol sizes, and types of paper or inks. For library systems, standards will have to be followed for proper bibliographic control. All but a few applications will require 100-percent checking of the original microfilm, especially the systems that involve the trouble and cost of preparing unit-records, such as aperture or Actifilm cards. This is particularly true for systems that generate and distribute duplicate files for a large number of users.

The quality of the system is a feature which must initially be built into the sub-systems, and monitored and maintained by interested personnel within the system. Some of the checking can be concurrent with the required operations. For example, equipment is available that will permit a check of an aperture card film image to be made at the same time that the film is being mounted, or at the same time that the identification is being keypunched from data on the film image. The equipments' technical specifications will depend to a great deal upon the nature and use of the original documents. For example, the filming of bank checks or legal documents will not require as much precision of detail as a system that must handle engineering drawings or aerial photographs. For any image system, every line, number, or notation on the original must be completely legible and non-ambiguous to the final user. As mentioned earlier, the viewing and printing specifications will usually dictate the filming and storage specifications. In particular, allowance must be made for the loss of resolution during each processing and reproduction step, and for the number of film generations that may be required. The resolution losses are inherent in every transfer of information from one medium to another, and will be caused by the imperfect nature of each of the parts of the processing chain such as the optics, mechanical structures, and the reproduction medium.

Publication in Microform

At least 47 organizations now publish books, reports, journals or other material in microfilm, Microcards, or some other microform. In many cases, the microform is a copy of a regular serial publication, such as the New York times, Aviation Week, Chemical Engineering, Time, Life, or Harvard Law Review, or a copy of some collection or publication that would be difficult for most libraries to acquire (e.g., the 4500-page transcript of the trial of Adolph Eichmann, or the 100,000 pages of the Annals of the French Chamber of Deputies). There is at least one periodical, Wildlife Diseases, that is published exclusively in Microcard form (by the American Institute of Biological Sciences) with a separate card for each article. Many non-periodical publications, such as patents, government reports, and dissertations are available in microform. One organization, for example, has microfilmed over 7000 dissertations to date, and now regularly microfilms the dissertations of more than 100 colleges and universities. The same organization has also made arrangements with the publishers of approximately 1200 journals to provide microfilm copies of their publication. It is estimated that 90 per cent of the current U.S. newspapers are available in microfilm form.[26] The cost of material published in Microcard and roll microfilm form is about 0.5 and 0.25 cent per original page, respectively. Approximately 30,000 different titles are currently available in microform.[27] Some of these works are over 10,000 pages long. One company has recently published a collection of more than 60,000 pages of manufacturers' catalog data in microform for engineers and other manufacturing personnel who are interested in locating the venders of specific types of parts or equipment. The catalog pages are reduced and printed 72 pages to each 4-by-6-inch acetate card. The pages are filed by company name and cross-referenced with a manual index.

There seem to be four main situations where a good case can be made for the use of microform publication: (1) where the cost of microform is

[26] Power, E., "Microfilm as a Library Tool," *Special Libraries*, Vol. 51, No. 2, pp. 62–64 (February 1960).

[27] Diaz, A. J., editor, *Subject Guide to Microforms in Print, 1962–1963,* Microcard Editions, Washington, D.C. (October 1962), 88 pp. This is a guide to the works of 47 publishers of information in microform.

Tilton, E. M., *A Union List of Publications in Opaque Microforms* (Scarecrow Press, New York, 1959). This is a list of some 3200 items of 23 publishers.

Power, E., "Microfilm—The Versatile Academic Tool," a section in *Microtezts as Media for Publication*, pp. 10–26 (Hertfordshire County Council, Hatfield, Herts, England, 1960).

Rice, S., "Publishing in the Microforms," *Proceedings of the 11th Annual Conventwn of the National Microfilm Association*, pp. 271–276 (National Microfilm Association, Annapolis, Maryland, 1962).

Simonton, W., "Library Handling of Microfilm Publication," *Proceedings of the 11th Annual Conventwn of the National Microfilm Association*, pp. 277–282 (National Microfilm Association, Annapolis, Maryland, 1962).

much less than the binding and storage costs for the original material, or where storage space is extremely scarce; (2) where the initial distribution is small enough that microform is a less expensive medium for primary publication; **(3)** where the only available copy already exists in microform; (4) where a compact file is desired for individual and personal use. Each of these situations is discussed in more detail in the following paragraphs.

It is estimated that it costs a library about 25 cents a year to store each book and about 50 cents a year to store each volume of bound periodicals, plus about 5 dollars per volume to do the binding. The cost of a microfilm copy may well be about the same as the cost of binding the periodicals.[28] To install microfilm copies without interrupting the library service, and with the least inconvenience to the users, libraries often subscribe to the original paper issue of the journal as well as to the complete microfilm volume of that journal. At the end of the volume year, a microfilm record of the entire year's issues is received and filed, while the original paper copies are retained only until their period of greatest use is over, until they are worn out, or until there is no more space available. At that time, the paper copies are thrown out and the microfilm copies substituted. The microfilm dealers' arrangements with the original publishers are such that microfilm sales are usually limited to subscribers of the original paper edition. Microfilm duplicates are especially useful for newspaper collections.

For limited distributions, it may be far less expensive to publish and distribute in microform than in standard book form. Whereas a minimum edition for a book publisher may be anywhere from 500 to 5000 copies, the minimum edition for books in Microcard form is 15 copies. The journal Wildlife Diseases, published initially for some 125 subscribers, is an example of microform publication for a limited audience.

Many books are published in limited editions, and become unavailable when out of print. A copy of an out-of-print book can be put into microfilm form for distribution, and the microfilm used later as a medium for xerographic printing equipment for on-demand printing, since it is less expensive to make xerographic prints from microfilm than from the original book. Xerox copies of out-of-print books of over 50 publishers can be obtained through microfilm and xerographic printing techniques by some of the microfilm publishers and service bureau~ Single copies of these Xerox books are commercially available on regular book paper for $4\frac{1}{2}$ cents per page, including binding. For a larger number of reprints, paper master plates could be prepared from the microfilm so that the final cost would range from 1.5 to 2.5 cents per page for 100 to 50 copies. The use of microfilm techniques is also being considered for the on-demand printing printing problem of the U.S. Patent Office, which now furnishes copies of 25,000 patents per day for public sale and internal use.[30]

In some cases, microform copies permit large files to be made available for small offices or individual users, and allow a large amount of data to be conveniently and neatly assembled in one location. Microform copies of many law books and statutes are currently available for the practicing attorney's private office. The State of California Appellate Court Reports, for example, have been distributed in Actifilm form (see Fig. 9-14), with 224 pages on each 5-by-8-inch Actifilm card. The sheets can be duplicated and disseminated at a cost of about 14 cents each. As another example, the publication of the meteorological data produced by the International Geophysical Year is expected to provide subscribers with 24 trays of 3-by-5-inch Microcards, as opposed to some 750 feet of shelving that would be necessary if this information were published in conventional letterpress form. The Microcard cost to each subscriber will be about $5000, as contrasted with the $60,000 that would have been required for conventional letterpress printing.[31] Some experimental work is being done by the Bell Laboratories to provide a microform copy of large telephone directories for use by the telephone companies' "Information" operators. The New York

[28] Power, E., "Microfilm as a Substitute for Binding," *American Documentation,* Vol. 2, No. 1, p. 36 (January 1951).

Meals, F. L., and W. T. Johnson, "We Chose Microfilm," *College and Research Libraries,* Vol. 21, No. 3, pp. 223–228, May 1960.

[29] Power, E., "O-P Books: A Library Breakthrough," *American Documentation,* Vol. 9, No. 4, pp. 273–276 (October 1958).

[30] Urbach, P., "A Future Microsystem for the U.S. Patent Office," *Proceedings of the 11th Annual Convention of the National Microfilm Association,* pp. 153–164 (National Microfilm Association, Annapolis, Maryland, 1962).

[31] Clapp, V., "Journal Publication in Microform," *Science,* Vol. 127, No. 3307, p. 1145 (May 16, 1958).

City operator, for example, has a working station with an alphabetical and street-address directory of some 4000 pages, which weighs about 25 pounds. Some experiments were made to replace the bulky file with a small deck of 6-by-8-inch micro-opaque cards. The field trials tested formats for 28 directory pages per card, and for 60 directory pages per card.[32]

Another example of the use of microform for the dissemination of technical information is the publication and distribution of technical reports in Microcard form by the U.S. Atomic Energy Commission. An average of 300 copies of each report are produced and distributed to a regular mailing list.[33] As of January 1962, over 20,000,000 Microcards have been distributed by the AEC.

General Business Records

The microfilm systems used for business records are intended primarily to save storage space. The emphasis is on a reduction in file size, rather than on rapid searching of the files or high-volume reproduction of the file contents. Most businesses utilize more roll microfilm than any other microform. Roll microfilm is much cheaper than aperture cards or the forms of other techniques, and the relatively slow access time and the awkwardness of the reproduction facilities are not serious obstacles. If high-speed roll film searching is required, such additional hardware as the Benson-Lehner FLIP equipment may be used. However, automatic file-searching equipment requires some digital coding on the film, and the manual indexing, along with the equipment for putting the coding pattern on the film, will add to the cost of the system. Some indexing is currently provided on manual roll systems by including target frames or title and separator frames with the other film images.

For the inactive storage of business or other records, it may be cheaper to keep the original records in some relatively inexpensive warehouse location and locate the clerical records manually, rather than microfilming the entire collection. Some consulting firms suggest that where a savings in storage cost is the only objective of microfilming, records to be retained 10 years or less should not be microfilmed.[34]

Some representative examples of the use of microfilm for applications other than those mentioned in the previous sections are described below.

Medical Records. Some hospitals have converted their patient files to microfilm systems. For this type of application, it must be relatively easy to add a new entry into a patient's record. This may easily be achieved with such techniques as film jackets. One particular scheme which is applicable to this type of file maintenance problem is opaque Microtape, described earlier in this chapter.[35] One of the large teaching hospitals in this country has filed its case histories in film jackets for use by the doctors and interns. Nearly 500,000 jackets have been prepared, and over 50 film viewers are used in this installation.

Personnel Records. A large industrial organization has replaced some 12,000 file folders of personnel records with 16mm film jackets. The use of the film jackets allows the individual records to be updated with additional information by the insertion of new frames of 16mm film into the film jacket.

Storage for On-Demand Printing. The Armed Services Technical Information Agency (ASTIA) has one of the largest collections of scientific and technical reports in the world. There are approximately 750,000 separate reports in the collection, received from more than 10,000 different corporate authors. ASTIA receives several hundred thousand new reports annually and selects approximately 30,000 of these for entry into the file system. For reproduction and distribution services, full-size Xerox copies are generated from short microfilm strips or records of the original document. There is a current collection of 750,000 such strips, and the strips for

[32] *Osten-Sacken, I. C., "The 'Information' Problem," Bell Laboratories Record, Vol. 37, No. 5, pp. 162168 (May 1959).*

[33] Diaz, *A. J., "Micro-Opaque Publishing," Proceedings of the 9th Annual Convention of the National Microfilm Association. DD. 218–226* (1960).

[34] *Retention and Preservation of Records (With Destruction Schedules), 6th ed., p. 9 (Record Controls, Inc., Chicago, Illinois,* 1961).

[35] Besserer, *R. T., "1,500,000 Medical Records in One Cabinet," Systems, Vol. 20, No. 2, p. 5 (March* 1956).

Whittaker, E. R., "The Application of the Microtape System to the Operation of a Prepaid Medical Care Plan (Blue Shield)," Proceedings of the 9th Annual Convention of the National Microfilm Association, pp. 179–181 (National Microfilm Association, Annapolis, Maryland, 1960).

Parche, M. C., "This Works for Us . . . Microcards in an Edition of One," Special Libraries, Vol. 50, No. 1, pp. 36–37 *(January 1959).*

approximately **1200** different documents must be assembled for printing every day.

Credit Reports. Dun & Bradstreet writes approximately 5,000,000 business credit reports each year. Approximately half of these reports become inactive, because the business has dissolved, or for some other reason. For nearly 100 years, a carbon copy of each inactive report was filed. This system gradually grew too bulky and awkward to use, and a changeover to microfilm technique was initiated. The reports are currently reduced 40:1 on 16mm roll film, and are used with reference viewers and reproduction facilities. The officereceives approximately 200 calls per day*requesting information from these files.

Legal Records. Births, marriages, deaths, and all the vital records pertinent to the entire population of one state in the U.S. are mounted on 35mm aperture cards.

All articles of incorporation filed in another state have been microfilmed and filed in film jackets and indexed for active use. Over 100,000 jackets have already been used for filing the back records.

A New England county uses aperture cards and film jackets for the filing, searching, and reproduction of all of its land transaction records. Approximately 150,000 documents per year are filmed and mounted.

The Survey, Titles, Records, and Drafting Branch of the New Brunswick Department of Lands and Mines has custody of approximately 10,000 original grants, land plans, and associated correspondence for the Province of New Brunswick. The documents date from the 1700's to the present, and must be examined when land titles are being searched. The documents have been put on rolls of 35mm film, and viewers are now used for title searches.

A large Federal agency utilizes aperture cards for the filing and reference of nearly 4,000,000 documents comprising the land grants, oil maps, and mining maps that cover the United States.

Motor vehicle registration data for the entire state of Connecticut are contained in a file of 16mm aperture cards. Two sets of data cards and one aperture card are generated for each title. The three sets of cards (750,000 vehicles per set) are contained in approximately 30 file cabinets. One data card is filed alphabetically by vehicle owner, the second data card is filed in order by vehicle identification number, and the aperture card is filed by title number. Microfilm printing equipment is used to make 2000 to 3000 title certificates daily.

Library Operations. Microfilm has been used to help keep track of articles charged out of a library. "Overdue" notices are made by clerks who check a transaction record on microfilm (photo of book card and borrower's library card). Each roll of film is dated and numbered to agree with numbers on the date-due cards placed in the book at the time of circulation.[36]

Microfilm copies (16mm) of the lists of holdings of the larger Canadian libraries were used to help compile a Union Catalogue for the National Library of Canada.[37]

One other possible application is the use of microfilm copies for interlibrary loan--either for direct loan, or for preparation of photocopies.[38] This method may be too expensive for most loan material, but it may be a convenient way to permit the loan of unique and rare documents.

Engineering Drawing Systems

Basic Systems. The engineering drawing systems warrant some discussion here because they currently appear to be one of the most active areas for microfilm storage and retrieval systems. All but a few of these systems use a unit-record concept, and the storage medium in most cases is the aperture card. Nearly all of the equipment developed to use aperture cards has been aimed at the engineering drawing application. These engineering drawings are generally in standard sizes which are multiples of 8½ by 11, or 9 by 12 inches, up to a usual limit of 36 by 48 inches. Larger drawings are used, but they are usually kept in roll form. The drawings are generally in pencil or ink on tracing cloth or transparent vellum. The full-size drawings are usually kept in files of flat drawers. A file

[36] **Anon., "Library Scores with Motto of Better Service,"** *Systems Management,* **Vol. 2, No. 4, p. 18 (October 1961).**

Kingery, R. E., "Copying Methods as Applied to Library Operations," *Library Trends,* **Vol. 8, No. 3, pp. 407–413 (January 1960).**

[37] **Patterson, F. E., "A New Reference Librarian Looks at the National Library,"** *Ontario Library Review,* **Vol.** *42,* **No. 1, pp. 35–39 (February 1958).**

[38] **Giinther, A., "Microphotography in the Library,"** *UNESCO Bulletin for Libraries,* **Vol. 16, No. 1, pp. 1–22 (January–February 1962).**

cabinet with ten drawers can hold up to 1500 drawings. The original drawing must be easily accessible so that it can be altered to show the changes in the design, tolerances, or materials. Changes are usually handled by erasing part of the original drawing, redrawing the appropriate portion, and noting in a space provided the change that was made. Most drawings will have at least 2 or 3 changes on them, and some may have as many as 20 or 30 changes. This circumstance, and the usual requirement that all interested parties automatically receive the newest prints, are common requirements of the storage and retrieval system.

An example of the magnitude of the drawing problem is the fact that the Department of Defense's annual expenditure for drawings and reproductions is currently over $2,000,000,000.[39] The files of engineering drawings for the military services contain some 50,000,000 drawings, with an input rate of 6,000,000 new or revised drawings per year. An estimated 1,000,000,000 prints are generated and distributed each year to government agencies and contractors. Some individual plants or centers produce as many as 3,000,000 prints per year. It might be noted that over 28,000 drawings are required to describe a new airframe for such a plane as the B-47 jet bomber. This storage problem can be reduced somewhat if the drawings are kept on film instead of full-size paper stock, and because of this, some government agencies actually require the contractor to furnish final prints in microfilm form. An example of the space reduction possible is given by the U.S. Army Signal Supply Agency, which converted the contents of 600 vertical files into aperture cards, which were then stored in 9 mechanized card files containing approximately 250,000 aperture cards apiece.

Another example of a large drawing file is in the East Pittsburgh Westinghouse plant, where 400 draftsmen are actively engaged in creating new or revised drawings.[40] More than 30,000 new drawings are created, and more than 54,000 drawings are modified, each year. More than 3000 requests for a total of about 16,000 prints are processed daily, and this reproduction process consumes 20,000,000 square feet of paper each year.

In one other example, the Bell Telephone System expects to save 10,000,000 dollars a year in distribution costs by sending microfilm copies of engineering drawings in place of blueprints to their operating telephone companies, Western Electric offices, and the Bell System Equipment Engineering Organization.[41]

The Use* of *Aperture Cards. Much of the microfilm production in this country is on 35mm film for microfilming engineering drawing systems. A little 105mm film is also used. Recent Department of Defense efforts to standardize the engineering documentation of the military and contractor establishments working on defense contracts have resulted in a system under which a great many manufacturers will be required to submit 35mm microfilm copies of their drawings rather than the original drawings.[42] This Department of Defense directive has further stimulated many contractors to establish their own microfilm systems. In many cases, significant cost reductions can be achieved for the manufacturer by the introduction of such a system.

The Department of Defense standards require a minimum camera resolution of 120 lines per mm throughout the optical field when installed and ready for use. The standards also require a resolu-

[39] Dunn, J. J., "Control, Administer, Automate . . . Reproduction of Engineering Drawings in the Department of Defense," *Filmsort Facts,* Vol. 2, No. 8 (December 1959).

[40] New, T. H., "Making the Decision to Automate Engineering Drawings," *Filmsort Facts,* Vol. 1, No. 12 (November 1958).

McMahon, G. J., "The Use of the Filmsort Card in Industry. The American Documentation Institute Meeting in Miniature," *American Documentation,* Vol. 12, No. 3, pp. 219–221 (July 1961).

[41] Locke, W. J., and C. E. Nelson, "Ten Million Dollars a Year in Savings," *Industrial Photography,* April 1962, pp. 36-39.

Locke, W. J., "Microfilm Pushes Drawings Aside," *Product Engineering,* October 27, 1958, pp. 44-47.

[42] Hutchinson, W. S., "Department of Defense Engineering Data Micro-Reproduction System," *Proceedings of the 9th Annual Convention of the National Microfilm Association,* pp. 80–82 (National Microfilm Association, Annapolis, Maryland, 1960).

Wooley, O. W., "Development and Implementation of Department of Defense Engineering Data Micro-Reproduction System, EDMS-0009," *Proceedings of the 9th Annual Convention of the National Microfilm Association,* pp. 84–87.

Borden, F. R., "Proposed Military Standard and Specifications on Engineering Data Tabulating and Aperture Cards," *Proceedings of the 9th Annual Convention of the National Microfilm Association,* pp. 89–91.

Anon., "Requirements for Microfilming Engineering Documents on 35-mm Film—A Proposed Military Specification," *The National Micro-News,* No. 53, pp. 73–104 (August 1961).

tion of 200 lines per mm for silver and Diazo film, and 130 lines per mm for heat-developing film.[43] These resolutions are beyond the capabilities of most office microfilming systems.

Engineering organizations generally use microfilm and aperture cards only to the extent of mounting the master image or a copy of the master image to simplify their storage and re-filing. In a few instances, there are requirements for multiple sets of microfilm drawings for distribution; these requirements can sometimes be handled by using cameras that automatically generate two original rolls at the same time, or automatically produce multiple images of a drawing on the same reel. Two extensive drawing release systems have been described: (1) the Navy issues a complete set of engineering drawings for the Polaris missile system, in aperture card form, to each Polaris-carrying submarine; (2) the Chrysler Corporation Missile Division distributes 18 sets of aperture cards (i.e., 18 copies) for each of several thousand engineering drawings of the Redstone missile, and 22 sets of aperture cards (i.e., 22 copies) for each of several thousand engineering drawings of the Jupiter missile.

In most systems, the original engineering drawings are retained in the plant so that they are readily accessible for revisions and reproductions. The first-generation-negative camera films are often sent to remote storage, primarily as security against the loss or damage of the original papers. The main use of the first-generation films is for security or disaster files, and intermediate prints are generated for distribution or aperture card mounting. In most cases, sets of second- or third-generation negative images for distribution purposes are produced in aperture card form. Figure 9-42 describes several of the more common methods for preparing negative image aperture cards from a master roll film for distribution. Copies can also be made of the distribution cards using Diazo or Kalfax techniques, which will actually result in a third- or fourth-generation image. At this stage in the system, the original image has been degraded by this chain of copy operations, and a good fourth-generation image will be possible only if the original filming was done with precision. This is part of the reason for insisting on high-quality filming. Whenever the original drawing is modified, it must be put through this filming and copying process again. There are many variations to this basic microfilm system, but the basic patterns are essentially the same.

For microfilm engineering drawing systems, most users prefer to retain the original camera negative in roll form. The roll form is preferred to the unitized form in order to retain complete file integrity by removing the possibility of losing or misplacing a single card. The original camera film (first-generation film) is preferred because the image is of a higher quality than those of subsequent copies.

There are differences of opinion among users regarding preferences for positive or negative film. The most common practice is to use a negative image both for the master roll which is to prepare multiple copies and for the distribution copies themselves. One reason for this preference is that viewing film records from a negative image is supposedly much easier on the eyes than is viewing from a positive image. Installations with a large number of viewers (e.g., the Social Security Administration with more than 100 viewing stations and operators) report that considerable eyestrain results from the background glare inherent with positive images, and that positive images are unacceptable. Another reason for the negative image preference is the claim that positive images produce poorer prints when used with many of the microfilm viewer-printers or xerographic printers.

No equipment yet on the market will handle the mechanized filing and retrieval of engineering drawings. If the microfilm images are on roll microfilm, there is no way to provide easy access to the particular frames of a roll without considerable wear and tear on the rolls.

Although most aperture card users mount images on standard punched card stock, very few of them exploit this feature by using tab card equipment with the system. The aperture card is designed to be used with any of the punched, card tabulating equipment (sorters, reproducers, collators, printers), but in nearly all cases, the user will not allow the image cards to be passed through any equipment after the film has been mounted. This is primarily because of the possibility of scratching the film image, or mangling the entire card during a card

43 White, R. A., "Specifications for Raw Stock Film for the EDMS-0009 Program," ***Proceedings of the 9th Annual Convention of the National Microfilm Association,*** pp. 93–96 (National Microfilm Association, Annapolis, Maryland, 1960).

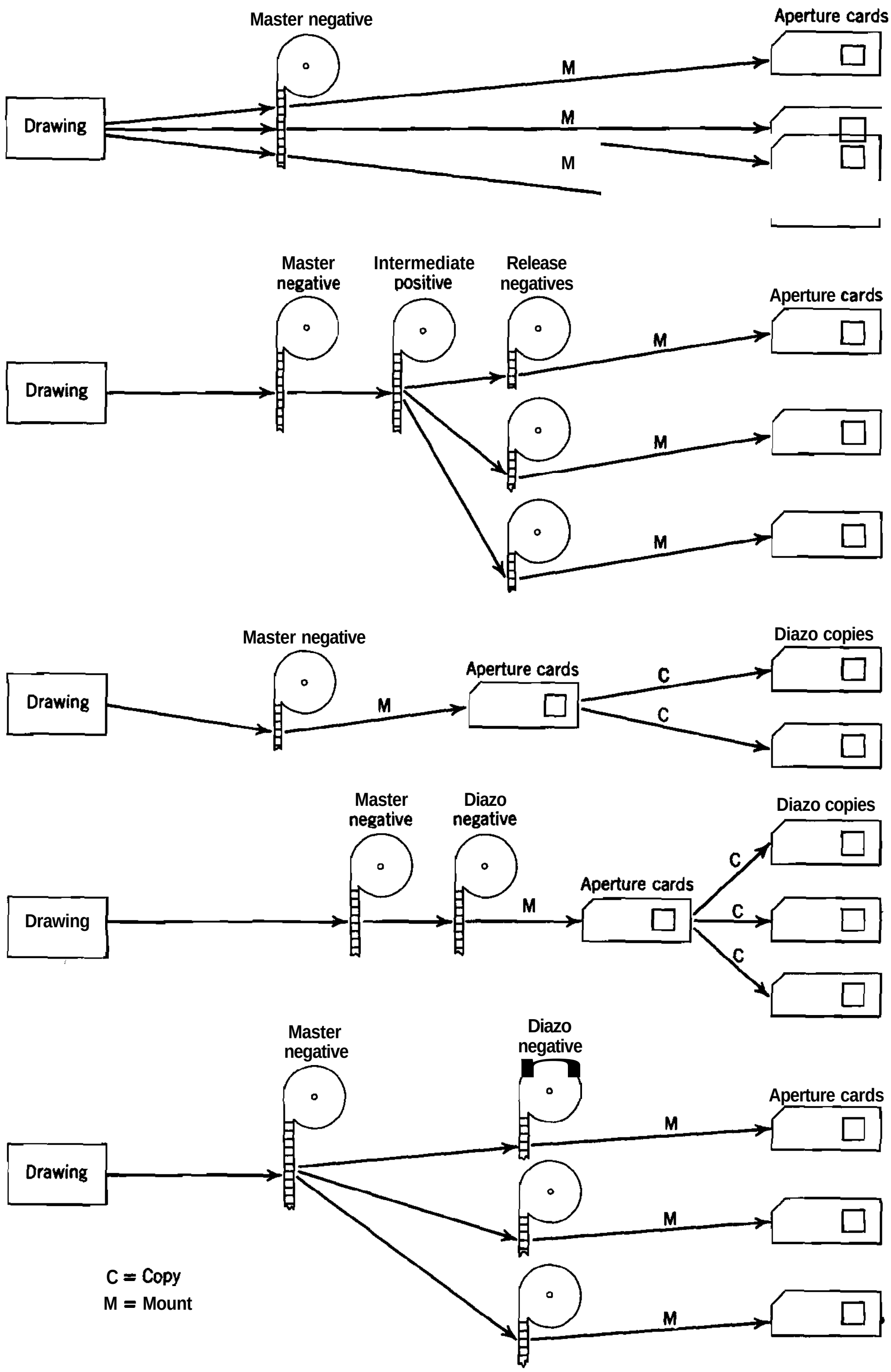


FIG. 9-42 Common methods for obtaining multiple copies of negative microfilm images in aperture card form.

jam.[44] Some users have found ways to use the cards in the punched card equipment before the film is mounted, in order to perform such functions as duplicating, sorting, listing, collating, and interpreting. Another problem with machine handling of the aperture cards is that each pass through the machine, especially for high-speed equipment, inflicts some wear on the card stock, and will eventually fray the card edges in such a manner that the cards can no longer be handled by the card equipment. For high-speed sorting or collating equipment, a deck of aperture cards might be unacceptable to the machine after several hundred passes. Slower machine speeds will increase the life of the card, but not to any significant degree. It is hoped that card equipment will eventually be designed to eliminate wear, but until then, no extensive or permanent information system should be designed around the concept of manipulating the aperture cards in punched card equipment.

One way to bypass some of these card-handling problems is to generate a punched card copy of the data on the aperture card, and manipulate this card instead of the aperture card. Data cards can easily be reproduced if the cards are mangled or damaged by the machine. This would permit the use of tab equipment to search the file contents, generate lists or reports, insert new material into the appropriate file location, and check the file sequence or order. Data cards can also be used to reproduce some of the indexing information on the aperture card before the film is mounted—thus labeling the top edge of the aperture card without extensive manual or clerical effort. However, other approaches have been used.[45] Many companies keypunch directly onto the aperture cards, and check the keypunching by proofreading the card against a printed list. In some cases the punching is done directly from the original or filmed drawing, and the card proofread against the drawing. In many cases, a semi-automatic optical mounter is used to permit an operator to check the filmed drawing against the card title before the film is inserted into the cards. A quality control check on each film image can also be made at this point. For convenience, the aperture cards are usually printed with a format which defines the fields on the card to suit the particular user, and makes the card labels easier to read. Different colors of card stock can also be used to indicate the film generation, source, or other characteristics.

Most drawing systems use similar identification or indexing schemes, which normally include

Drawing number
Drawing title
Issue number or change letter
Security classification
Distribution code
Originating engineering section
Drawing size or reduction ratio
Date
Related drawing numbers
Material
Process

Machine searching of the master punched card decks (not the aperture card decks) may allow users to find all the drawings related to one particular part or product, find all products made of a particular material or by a particular process, or list all drawings generated by a particular department or on a particular date.[46] Parts lists for the associated drawings can be kept on punched cards. It may often, for example, be desirable to know how many of each vender's items were needed for several related assemblies. In situations where reliability, cost, or reproduction design studies indicated that a particular component was to be avoided, it would be desirable to know which drawings or units were affected.

The Use of 70mm and 105mm Film. Many organizations are convinced that 105mm film is the best for engineering drawings. The primary argu-

[44] Weigers, J., "The Use of IBM Equipment with the Filmsort System," *The National Micro-News,* No. 53, pp. 5M8 (August 1961).

Carroll, H. L., "Use of Microfilm Aperture Cards in IBM Equipment," *Proceedings of the 9th Annual Convention of the National Microfilm Association,* pp. 118–122 (National Microfilm Association, Annapolis, Maryland, 1960).

[45] Banar, P. D., and C. E. Nelson, "Comparative Methods of Providing Information for Keypunching Aperture Cards," *Proceedings of the 9th Annual Convention of the National Microfilm Association,* pp. 137–139 (National Microfilm Association, Annapolis, Maryland, 1960).

McMahon, G. T., "Shortcuts to Automation," *Proceedings of the 10th Annual Convention of the National Microfilm Association,* pp. 48–54 (National Microfilm Association, Annapolis, Maryland, 1961).

[46] Anon., "Microfilm Speeds Data Distribution for Army Signal Corps," *ODR Reproductions Review,* Vol. 10, No. 5, p. 12 (May 1960).

ment is that the larger the negative, the less the drawing has to be reduced and re-enlarged, and thus the less risk there is of loss of detail and distortion. Systems with minor image reductions can tolerate more variation in the production parameters and do not need as close control as a 35mm system. An additional advantage is that the 105mm negative (approximately 4 by 6 inches) can usually be hand-sorted, filed, and read in a fair amount of detail without optical aids.

Many engineering organizations use this system. The Atomic Energy Commission, for example, has photographed over 15,000 drawings of nuclear processes and equipment, and makes 105mm prints available to interested schools and industries. The U.S. Corps of Engineers distributes drawings to regional offices in 105mm form. The 70mm systems are also in common use. The Babcock & Wilcox Co. has recently converted a 1,500,000-engineering-drawing file to a 70mm system.[47]

The Use of Continuous Image Film. The CIM-card was developed to satisfy a particular problem of the Customer Service Department of Douglas Aircraft.[48] This company had established service offices in several cities in order to provide regional information centers for their field representatives and the airplane customers. Each office had to be supplied with a complete set of prints for the DC-6 and the DC-7 airplane. This amounted to a total of approximately 45,000 prints per office, which required 75 file cabinets for storage. The file size was due to double when prints became available for the DC-8. This situation initiated the development of the CIM-card and its associated equipment. The 75 file cabinets were subsequently replaced by one file cabinet of CIM-cards and a viewer-printer machine.

COSTS

Equipment Costs

There is a very wide range of equipment costs for the image systems, depending primarily upon the degree of mechanization and sophistication required. Representative costs of much of the microfilm and image-handling equipment are given in Tables 9-1 and 9-2.

TABLE 9-1

Representative Costs for Common Microfilm Equipment

Equipment	Approximate Purchase Cost
Cameras for general business recording	$ 350 to 2,700
Cameras for precision filming	3,000 to 8,000
Viewers	250 to 1,000
Viewer-printers	650 to 900
Contact printers (roll-to-roll)	1,800 to 3,100
Aperture card copier (slow-speed)	750
Aperture card copier (high-speed)	25,000
Hard copy printer (high-speed printing from aperture cards)	8,000 to 160,000
Aperture card mounter (manual)	700
Aperture card mounter (high-speed)	33,000
Microfilm storage cabinets (5-drawer)	200 to 260
Microfilm storage cabinets (10-drawer)	250 to 440

TABLE 9-2

Representative Costs for Mechanized Image Systems

Equipment	Approximate Purchase Cost
Benson-Lehner FLIP	$ 40,000 to 50,000
FMA Filesearch	143,000 to 157,000
Ferranti-Packard Rapid Access Look-Up System	17,000
Eastman Kodak Minicard	over 1,000,000
Filmorex system	25,000
IBM WALNUT system	500,000 to 1,000,000
AVCO Corp. VERAC 903	over 100,000
Video tape system	500,000 to 1,500,000
Recordak Lodestar with counting accessory	4,600
CRIS (Information for Industry)	33,500
CRIS scroll preparation unit	10,000
Magnavox MEDIA Camera	8,500
Magnavox MEDIA Film Cutter	1,500
Magnavox MEDIA Selector-Reproducer	25,500

Material Costs

The basic materials used in image systems are the film stocks, paper, and jackets or mounting frames. Representative costs are given in Table 9-3.

[47] Swarmer, K. E., "70mm Microfilm Program for 1,500,000 Engineering Drawings," *Industrial Photography,* Vol. 5, No. 8, pp. 24–25 (August 1956).
[48] Anon., "Jet Age Demands Miniaturization," *Industrial Photography,* Vol. 10, No. 1, p. 92 (January 1961).

TABLE 9-3

Representative Material Costs for Image Systems

Item	Approximate Cost
Silver negative film (35mm)	$ 5.70 per 100-foot roll
Silver negative film (35mm) including processing	6.50 per 100-foot roll
Silver negative film (16mm)	3.10 per 100-foot roll
Silver negative film (16mm) including processing	4.90 per 109-foot roll
Silver negative film (extremely high resolution Lippman emulsion)	0.50 per 3-by-5-inch sheet
Silver positive film (35mm)	3.60 per 100-foot roll
Kalfax film (35mm)	4.00 per 100-foot roll
Diazo microfilm (35mm)	1.60 per 100-foot roll
Blank aperture cards (35mm)	40.00 per thousand
Diazo aperture cards (35mm Duplicards)	45.00 per thousand
Kalfax aperture cards (35mm KalKards)	53.00 per thousand
Kalfax sheet film (tab card size)	119.00 per thousand
Actifilm sheet film (tab card size)	112.00 per thousand
Actifilm sheet film (5-by-8-inch)	182.00 per thousand
Actifilm sheet film (3-by-5-inch)	71.00 per thousand
Copy paper and other expendables for viewer-printers	0.05 to 0.12 per 9-by-12-inch page up to 0.20 per 18-by-24-inch page
Copy paper and other expendables for high-speed xerographic or electrostatic printers	0.05 to 0.02 per 9-by-12-inch page
Opaque microstrips with adhesive backing (16mm)	9.50 per 100-foot roll
Opaque microstrips with adhesive backing (35mm)	13.50 per 100-foot roll

Process Costs

The approximate costs that might be charged by service bureaus for several of the more common microfilm operations, such as film developing, film copying, and aperture card preparation, are given in Table 9-4. These are costs that commercial microfilm bureaus might charge. For microfilming operations that are to be done within an organization, the operating speeds shown in Table 9-5 may serve as useful rule-of-thumb guides to the amount of equipment required. In general, actual operating or processing costs are relatively difficult to obtain. However, a' few publications have made reference to the cost considerations.[49] A few cost estimates are included here to aid in making an initial estimate of the cost of a proposed microfilming program. They should not be considered as standard times or costs.

The following rules of thumb may also be useful to obtain preliminary estimates of equipment requirements:

1. Aperture cards may be filed with a packing density of 105 cards per file inch. One standard tab card file cabinet will hold 65,000 tab-size aperture cards (3250 per drawer). A motorized tub file unit may hold as many as 180,000 to 350,000 tab cards.
2. Three-by-five-inch Microcards (one-sided) can store up to 80 pages and be shelved 85 cards to the inch; two-sided Microcards can store up to 160 pages and be shelved 65 cards to the inch.
3. The filming of engineering drawings or other large images with 35mm film will result in about 550 to 750 images per 100-foot roll. If this film is to be mounted in aperture cards, then a standard film advance of 2 inches per image is used, resulting in about 500 images per 100-foot roll.
4. The filming of page-size business records or documents with 100-foot rolls of 16mm film will result in up to 1600 images per roll for a 16:1 reduction, about 2400 images per roll for a 20:1 reduction, and up to 7200 images per roll for a 34:1 reduction. Over 28,000 3-by-5-inch cards can be stored on a single 100-foot roll of 16mm film with a 40:1 reduction.
5. One hundred rolls of 16mm film can be stored in one drawer of a microfilm file cabinet.
6. Remington Rand Snap Jack aperture cards may be filed 62 cards to the inch.

49 Heilprin, L. B., "The Economics of 'On Demand' Library Copying," ***Proceedings of the 11th Annual Convention of the National Microfilm Association,*** pp. 311–339 (National Microfilm Association, Annapolis, Maryland, 1962).

Luther, F., "The Economics of Microfilming: A Comparison Between Contracted Services and a Do-It-Yourself Program," ***Proceedings of the 10th Annual Convention of the National Microfilm Association,*** pp. 33–40 (1961).

Banar, P. D., and C. E. Nelson (footnote 45).

Forbes, E. J., and D. P. Waite, ***Costs and Material Handling Problems in Miniaturizing 100,000 Volumes of Bound Periodicals, a*** report to the Council on Library Resources (Forbes & Waite, Lexington, Massachusetts, August 1961).

TABLE 9-4

Representative Microfilm Processing Costs

Operation	Approximate Cost
Total Cost (including labor and materials)	
Microfilm books on 35mm film	3½ to 4 cents per page
Filming of library-type material for general reference and subsequent xerographic printing	2 cents per page
High-quality filming of library-type material	7 to 10 cents per page
Film engineering drawings, process and mount the film in aperture cost	30 cents per drawing
Filming of engineering drawings to exact Department of Defense specifications (120-lines-per-mm resolution, etc.)	24 cents per image for sheet drawings and 38 cents per image for roll drawings (A through E size) to produce images on rolls of silver negative film If some of the specifications can be relaxed, the cost will go down to 15 and 25 cents per image for sheet and roll film, respectively.
Photographing of the contents of one 4-drawer filing cabinet of business records	$ 90.00 to 110.00
Copy roll film onto Kalfax aperture cards	100.00 per thousand cards
Copy roll film onto sheet microfilm	30.00 per thousand cards
Duplicate roll film (silver to Kalfax)	7.50 per 100-foot roll
Duplicate roll film (silver to Diazo)	7.50 per 100-foot roll
Aperture card mounting	0.03 per card
Duplicate aperture cards	70.00 to 150.00 per thousand cards
Processing Cost Only (excluding material costs)	
Filming of engineering drawings	$ 70.00 to 150.00 per thousand drawings
Process silver film originals (35mm)	1.90 per 100-foot roll
Duplicate roll film (silver to silver)	3.50 per 100-foot roll
Duplicate roll film (Kalfax to silver)	3.50 per 100-foot roll
Duplicate roll film (silver to Kalfax)	2.50 per 100-foot roll
Duplicate roll film (Kalfax to Kalfax)	2.50 per 100-foot roll
Duplicate roll film (silver to Diazo)	2.50 per 100-foot roll
Mount aperture cards	63.00 per thousand cards some say 15.00 per thousand cards some say 100.00 per thousand cards
Make duplicate aperture cards	36.00 per thousand cards
Copy roll film onto Kalfax sheet film (1 image per sheet)	30.00 per thousand cards
Copy roll film onto Actifilm cards (1 image per card)	30.00 per thousand cards
Copy roll film onto Kalfax aperture cards	30.00 per thousand cards
Copy roll film onto Diazo aperture cards	30.00 per thousand cards

7. Kard-a-Film acetate jackets may be filed 36 to 44 cards to the inch, depending on the type of card used.

8. Approximately 12,000 pieces of 105mm sheet film can be stored in a standard cabinet for 5-by-8 cards.

9. Finally, 384 boxed reels of 35mm film can be stored in a standard 6-drawer microfilm filing cabinet, and 576 rolls of 35mm film can be stored in a standard 9-drawer microfilm filing cabinet.

Additional References

Bagg, T., and M. E. Stevens, *Information* Selection Systems Retrieving Replica Copy: State-of-the-Art Report, National Bureau of Standards Technical Note 157 (August 1962).

TABLE 9-5

Representative Microfilm Processing Speeds

Operation	Effective Rate
Preparation of correspondence-type copy for filming (removing staples, etc.)	2 to 3 times the filming labor
Filming of books or documents without special attention or equipment	2000 to 2500 pages per day
Filming of drawings or other large images	400 to 600 drawings per 8-hour day
Hand mounting of aperture cards	175 to 200 cards per hour
Hand copying of aperture cards (card-to-card)	up to 300 cards per hour

Ballou, H. W., *Guide to Microreproduction Equipment,* 2nd ed. (National Microfilm Association, Annapolis, Maryland, 1962). A factual guide, including illustrations and prices for all known microreproduction equipment made or sold within the United States.

Bishop, C., "Problems in the Production and Utilization of Microfiche," *American Documentation,* Vol. 12, No. 1, pp. 53-55 (January 1961).

Dugan, J. M., et al., "The Design of Reading Equipment for Library-Archival Utilization of Microforms," *Proceedings of the 10th Annual Convention of the National Microfilm Association,* pp. 94–101 (National Microfilm Association, Annapolis, Maryland, 1961).

Evans, A. D., *Engineering Data Processing System Design* (D. Van Nostrand Co., Princeton, New Jersey, 1960).

Grech, A. P., "The Development and Use of Microreproductions in the Libraries of the Legal Profession," *The National Micro-News,* No. 55, pp. 158–174 (December 1961).

Hawken, W. R., "Developments in Xerography: Copyflo, Electrostatic Prints, and O-P Books," *College and Research Libraries,* March 1959, pp. 111–117.

Hawken, W. R., *Production of Micro-Forms,* Vol. 5, Part I, of the series: *The State of the Library Art,* R. R. Shaw, editor (Rutgers University Press, New Brunswick, New Jersey, 1960).

Henkle, H. H., "Crerar Library Use of Microfilm in Science Information Service," *Proceedings of the 10th Annual Convention of the National Microfilm Association,* pp. 74–78 (1961).

Howerton, P. W., "A Unitized Microfilm Record System: A Case Study," *Proceedings of the 10th Annual Convention of the National Microfilm Association,* pp. 151–159 (1961).

Howerton, P. W., "The Case for Realism in Document Storage Systems," *Proceedings of the 11th Annual Convention of the National Microfilm Association,* pp. 21–25 (1962).

Hsiong, S. L., "The Application of Microfiche in Nanyang University Library, Singapore," *The National Micro-News,* No. 55, pp. 175–190 (December 1961).

International Federation for Documentation, "Manual on Document Reproduction and Selection," The Hague, 1953, 3 vols. (loose-leaf, kept up to date by supplements).

James, P., "A Photomagnetic System for Document and Information Retrieval," *American Documentation,* Vol. 10, No. 4, pp. 286–295 (October 1959).

Kiersky, L. J., "Bibliography on Reproduction of Documentary Information: 1955–1960," *Proceedings of the 10th Annual Convention of the National Microfilm Association* (1961). This is a compilation through 1960 of several bibliographies by L. J. Kiersky that have been published in *Special Libraries.* The 1961 bibliography was published in *Special Libraries,* Vol. 53, No. 3, pp. 135–140 (March 1962).

Korte, T. H., et al., *Microfilm Aperture Card System,* WADC Technical Report 59-746, Wright Air Development Center (June 1960), AD-232 960.

Lewis, C. M., and W. H. Offenhauser, *Microrecording—Industrial and Library Applications* (Interscience Publishers, New York, 1956).

Microtexts as Media for Publication: The Papers and Discussion of a Symposium Held at Hatfield Technical College on the 10th November 1969 (Hertfordshire County Council, Hatfield, Herts, England, 1960), 87 pp.

Parker, D. J., "New Tools for Micro-Recording and Reproduction," *Proceedings of the 9th Annual Convention of the National Microfilm Association,* pp. 158–164 (1960).

Patterson, H. D., "Microfilm System for Missile Documentation," *Proceedings of the 10th Annual Convention of the National Microfilm Association,* pp. 66-69 (1961).

Roudebush, L. A., "Industry Looks at the Department of Defense Microfilm Program," *Proceedings of the 10th Annual Convention of the National Microfilm Association,* pp. 70–73 (1961).

Scanlan, J. F., and T. G. Nanney, *The Active Uses of Microfilm,* a paperback book by Wolf Business Publications, Inc., April 1962.

Scott, P., "The Library Standards for Microfilm Committee," *Proceedings of the 10th Annual Convention of the National Microfilm Association,* pp. 103–106 (1961).

Skipper, J. E., editor, "Photoduplication in Libraries," The 10 papers in this special issue of *Library Trends,* Vol. 8, No. 3 (January 1960) are devoted to this topic, and include discussions of such things as microfilm, microtext, copying methods, and how these things affect the libraries.

Smith, G. L., "A Microfilm Library on Air Pollution," *Journal of Chemical Documentation,* Vol. 1, No. 3, pp. 11–15 (November 1961).

Soman, J. M., "Captive—Shop Precision Microfilm," *Proceedings of the 9th Annual Convention of the National Microfilm Association,* pp. 198–204 (1960).

Sophar, G. J., "The Ever Expanding Microform System—Reduction, Dissemination, Retrieval, Reproduction," ***Proceedings of the 9th Annual Convention of the National Microfilm Association,*** pp. 67–72 ***(1960).***

Stevens, G. W., ***Microphotography: Photography at Extreme Resolution*** (John Wiley and Sons, New York, 1957).

Stewart, J., and D. Hickey, ***Reading Devices for Micro-Images,*** Vol. 5, Part 2 of the series: ***The State of the Library Art,*** R. R. Shaw, editor (Rutgers University Press, New Brunswick, New Jersey, 1960).

Ten Eyck, H., ***Glossary of Terms Used in Microreproduction,*** rev. ed. (National Microfilm Association, Annapolis, Maryland, April 1962).