DOCUMENT RESUME

ED 051 844 LI 002 904

AUTHOR	Markuson, Barbara Evans, Ed.
TITLE	Libraries and Automation.
INSTITUTION	Library of Congress, Washington, D.C.
SPONS AGENCY	Council on Library Resources, Inc., Washington, D.C.; National Science Foundation, Washington, D.C.
PUB DATE	64
NOTE	265p.; Proceedings of the Conference on Libraries and Automation held at Airlie Foundation, Warrenton, Va. May 26-30, 1963
EDRS PRICE	EDRS Price MF-$0.65 HC-$9.87
DESCRIPTORS	*Automation, *Communication (Thought Transfer), Computer Graphics, Computers, *Conference Reports, Data Bases, Indexes (Locaters), Information Processing, Information Retrieval, Information Storage, *Libraries, *Library Networks, Microforms, Models, Systems Analysis

ABSTRACT

The arrangement of this publication follows, as nearly as possible, the actual conference program. The seven sections represent the topics selected by the planning committee as areas of concentration. Each section includes a state-of-the-art paper and related talks and discussion. The topics covered are: the library of the future, file organization and conversion, file storage and access of bibliographic information; graphic storage techniques and applications, microforms, output printing for library mechanization; library communications and networks, automation of library systems; library systems analysis; and mathematical models and system design. Biographical data and a list of the conference participants are included. (AB)

LIBRARIES and AUTOMATION

Proceedings of the Conference on Libraries and Automation held at Airlie Foundation, Warrenton, Virginia, May 26–30, 1963; under sponsorship of the

Library of Congress
National Science Foundation
Council on Library Resources, Inc.

Edited by
Barbara Evans Markuson

LIBRARY OF CONGRESS • WASHINGTON, D.C. • 1964

PREFACE

The arrangement of this publication follows, as nearly as possible, the actual conference program. The seven sections represent the seven topics selected by the planning committee as areas of concentration. In each section will be found a state-of-the-art paper and related talks and discussion, thus permitting the sections to be read independently. There is a slightly different format for Section I because the paper by Don R. Swanson was not distributed in advance but was delivered as the keynote address. Detailed tables of contents for each section are brought together at the front of the book, as a substitute, albeit inadequate, for an index.

The temptation to rewrite history has been scrupulously avoided: the conference sessions were tape recorded and are reported here with as much fidelity as possible. Space did not permit publication of the entire transcript of these sessions. Even so many who read these proceedings may feel that we have erred on the side of inclusiveness. Our criteria for inclusion of discussion was that it (1) relate directly to the paper being discussed, (2) give the reader additional insight into the subject under consideration, and (3) illustrate attitudes and uncertainties at this stage in the history of library mechanization. There were many personal sidelights at these meetings, allusions to which have generally been omitted because enjoyment of them depended largely on the immediacy of the situation. Occasionally we have included comments which cannot be defended on the basis of the criteria listed above, simply because these remarks retained, even in print, some of the flavor of the spirited and informal discussion which took place at Airlie.

It has been our aim to make this publication useful for subsequent reference. Therefore, all the bibliographies have been revised and presented in a uniform format and additional notes, headings, and comments have been provided to aid the reader. Notwithstanding the great desire on the part of both the sponsors and the participants to see tangible evidence of their accomplishments, conference proceedings demand the same editorial care as other publications. In fact, publication of such proceedings cannot be defended if they are to serve as expensive souvenirs for the participants, but only if they will be a worthwhile addition to the literature. We hope that these proceedings will be so regarded and that they faithfully reflect the intelligence, enthusiasm, and vigor which characterized the conference sessions.

BARBARA E. MARKUSON,
Library of Congress.

CONTENTS

SECTION II—Continued

CONFERENCE SESSION II

SECTION III—File Storage and Access

SECTION III—Continued

Page

CONFERENCE SESSION III

SECTION IV—Graphic Storage

SECTION IV—Continued

SECTION V—Output Printing

SECTION V—Continued

CONFERENCE SESSION V

SECTION VI—Library Communications Networks

SECTION VI—Continued

CONFERENCE SESSION VI

SECTION VII—The Automation of Library Systems

CONFERENCE SESSION VII

SECTION VII—Continued

APPENDIXES

LIST OF FIGURES

LIST OF FIGURES—Continued

INTRODUCTION

It is with some misgivings that librarians burden the world with the proceedings of yet another conference. In defense of this addition to the information explosion, we hope that the work reported here may help us toward the eventual solution of crucial problems in information handling.

A glance at almost any recent library journal will give evidence that librarians have already begun to look to mechanization as a solution to many of these problems. However, for some time there has been increasing concern in two of the major agencies supporting library research—the National Science Foundation and the Council on Library Resources, Inc.—about the number and quality of proposals which they receive. It is often obvious that those who designed these projects have not done their homework and are not only unfamiliar with the technical state of the art but with current library research. This overlapping and duplication of effort made it clear that something should be done to provide librarians with at least a broad acquaintance with the technology relating to library mechanization. This conference was conceived as a partial answer to this problem.

Librarians believe that this situation is not the concern of grant-making agencies only. Because we work in institutions which, by and large, do not make research and development funds available to the library, it behooves us, as members of a profession, to see that these scarce funds are spent wisely and that when projects do receive support, the results, whether positive or negative, contribute toward understanding and solving library problems.

On November 20, 1962, a group drawn from research libraries, foundations, and industry met at the Library of Congress to consider ways to bring about an improvement in this situation. There was a general reluctance to foster another conference on automation and general agreement that it might not be a suitable vehicle for accomplishing the desired objectives. For one thing, it would reach a limited audience. For another, it would be a one-time effort and its effect might not continue over a significant period of time. In spite of these caveats it was finally agreed that a working conference might provide the stimulus for librarians and technicians to destroy some of the stereotypes each group has of the other, to discover mutual problems, and to develop a common understanding of the goals of library mechanization and the ways in which each group could help in achieving these goals. The planning committee felt that by establishing this initial rapport, the conference might lead to other, more permanent, endeavors which would continue the work begun on the conference floor.

In order to minimize the well-known pitfalls of conference programming, the committee urged acceptance of the following ground rules:

1. Limitation of attendance to 100 people who were planning, or who had under way, mechanization projects.
2. Representation to follow a ratio of about two librarians to each technician.
3. Advance distribution of papers so conference sessions could be spent in active discussion.
4. Selection of library-oriented discussion leaders who could summarize the technical papers and relate them to library situations.
5. Concentration of the program on the major topics affecting library mechanization.
6. Publication of proceedings for the benefit of nonparticipants.

After working out a preliminary slate of topics, the committee suggested a panel of authors and discussion leaders and proposed that the Library of Congress assume responsibility for the administration of the conference and the editing of the proceedings. This publication gives evidence that the actual program adhered closely to the suggested format.

The sponsors did not expect that a meeting of this kind would result in a plan of action, a resolution, or even specific proposals. They did feel, however, that a list of recommendations based on an analysis of the conference discussion might prove helpful. Accordingly, they invited a small group to meet informally after the last session to consider such recommendations. Although a formal report did not result from this meeting, the following conclusions were reached:

1. Grant-making organizations, e.g. the National Science Foundation and the Council on Library Resources, Inc., should eliminate project reports and insist on publication in the open literature.

2. There is a need for a continuing, rapid, and informal reporting on library mechanization projects. This could be done by the Library of Congress, the Association of Research Libraries, the American Library Association, or some other group with an established publication program. At the present time there seems to be no need for establishing a separate organization for this purpose.

3. A variety of techniques (newsletters, journal articles, seminars, demonstration projects, and conferences) should be utilized to acquaint librarians with new developments in technology and information science. These media should promote communication within the library community and between librarians and other interested communities. The conference had demonstrated that librarians and industry people talk too much to their own groups and that cross-fertilization might be mutually advantageous.

4. In order to enlist the support of foundations, librarians must identify specific problems for study rather than propose broad, vague mechanization schemes. The following list represents some of the technical areas to which librarians and technicians might jointly direct their energies:

   File organization for storage and search
   Conver[illegible] techniques and methodology
   Compu[illegible]gram sharing
   Stand[illegible]ion of coding, etc.
   Man/machine interface
   Utilization of microforms in libraries
   Remote communication facilities
   Output printing techniques and requirements

   In addition, there are many library problems of a managerial nature which need further study. The following are, again, only examples:

   Development of manuals of procedures formalizing specific operations
   Development of better cost data for present operations
   Growth rate projections
   User acceptance of new techniques
   Copyright regulations affecting reproduction of materials
   Durability of various computer and microform storage media

5. Library schools need to train students in new techniques and methodologies; current programs should be evaluated to determine how they could be improved to prepare librarians and subject specialists in new information techniques.

6. Librarians must continue to standardize practices and to study changes that may be needed in descriptive cataloging rules, filing rules, and subject headings in order to prepare for mechanization.

During the planning stage, someone observed that a first conference on a subject often seems to fall short of the goal, but as time passes it becomes evident that much more was achieved than was immediately apparent, and eventually a first conference becomes the benchmark by which future meetings are measured. Should history be this kind to the Conference on Libraries and Automation, it will be due largely to those who gave their time and talents to its planning and execution—the planning committee, the authors of the state-of-the-art papers, the discussion leaders, and the arrangements committee. On behalf of the National Science Foundation, the Council on Library Resources, Inc., and the Library of Congress, I take this opportunity to acknowledge our indebtedness to these people and to Mrs. Barbara Markuson for her work in editing the preprints and the conference proceedings.

Just as important to the success of this conference were those for whom these plans were laid—the participants. These proceedings reflect the extent of their participation and the contributions which they made to every session. That we were able to gather such a distinguished group of busy people together on comparatively short notice is another evidence of the interest and enthusiasm the participants brought to this conference.

As you read these papers and the discussion, it may seem to those of you who have been librarians for many years that we have reached the end of an era. I counter this view with the fact that librarians historically have sought to improve their services by assimilating new technological advances and methods into library management—consider, for example, innovations like the book, the card catalog, the typewriter, the telephone, and the microfilm camera and reader. The quest for automation should be regarded as a continuation of our long tradition of change and improvement.

L. QUINCY MUMFORD,
Librarian of Congress.

CONFERENCE ORGANIZATION

Sponsors

National Science Foundation
Library of Congress
Council on Library Resources, Inc.

Planning Committee

Richard S. Angell, Library of Congress
Henry J. Dubester, Library of Congress
Bernard M. Fry, National Science Foundation
Robert M. Hayes, Advanced Information Systems, Inc.
Edward M. Heiliger, University of Illinois, Undergraduate Library, Chicago
Gilbert W. King, Itek Corporation
Foster Mohrhardt, National Agricultural Library
Frank B. Rogers, National Library of Medicine
Rutherford D. Rogers, Library of Congress
Melvin Voigt, University of California, San Diego

Session Chairmen

Burton W. Adkinson, National Science Foundation
Verner W. Clapp, Council on Library Resources, Inc.
L. Quincy Mumford, Library of Congress
Rutherford D. Rogers, Library of Congress

Arrangements Committee

Henry J. Dubester, Library of Congress
Barbara E. Markuson, Library of Congress
Marlene Morrisey, Library of Congress

Authors of State-of-the-Art Papers

Samuel Alexander, National Bureau of Standards
Lawrence H. Berul, Information Dynamics Corporation
Donald V. Black, Planning Research Corporation
J. W. Emling, Bell Telephone Laboratories, Inc.
James R. Harris, Bell Telephone Laboratories, Inc.
Gilbert W. King, Itek Corporation
Richard L. Libby, Itek Corporation
Harvey J. McMains, American Telephone and Telegraph Co.
Robert L. Patrick, Planning Research Corporation
F. Clayton Rose, National Bureau of Standards
David E. Sparks, Information Dynamics Corporation
Don R. Swanson, Graduate Library School, University of Chicago
David P. Waite, Information Dynamics Corporation

Discussion Leaders

Joseph Becker
Henry J. Dubester, Library of Congress
Foster Mohrhardt, National Agricultural Library
Frank B. Rogers, National Library of Medicine
Mortimer Taube, Documentation, Inc.
I. Albert Warheit, International Business Machines Corporation

735-898 O—64——2

Welcoming Address

L. QUINCY MUMFORD
Librarian of Congress

It is a pleasure to welcome each of you to this Conference on Libraries and Automation. The Library is indebted to the directors of the two foundations who have made this conference possible: Verner Clapp, the president of the Council on Library Resources, Inc., and Burton Adkinson, the head of the Office of Science Information Service, National Science Foundation. Both of them are committed deeply to the development of library and information services. It is our hope that some of the results of this conference can be useful to them in their future program formulation and in making the difficult decisions concerning areas of research and development that require encouragement and support.

We who are librarians have a range of problems which may be helped by automation. In my brief remarks this evening I will not try to detail the nature of the library profession and its needs. All librarians will probably agree that we have never been able to do all we have wanted to do in organizing our collections and serving them to the users of our libraries. All librarians see the emergence of new requirements and the possibilities of new services which are difficult to satisfy with the traditional levels of support afforded to libraries. Recent years have increased the hope that automation will do for libraries what they have been unable to do for themselves.

In this connection the Library of Congress, through the support of the Council on Library Resources, Inc., has been pursuing a survey of the possibilities of automating its information system. This survey has been conducted under the leadership of Dr. Gilbert King, who has had the assistance of a group of experts in computer technology. It is hoped that the report of this survey, which will be published later this year,[1] will be of some assistance to other institutions also. We at the Library of Congress know that we have not been alone in looking at the possibilities of automation and the use of computers. An increasing number of libraries have undertaken specific projects and activities using this modern technology or are planning to do so.

Our aim in preparing for this conference was to invite delegates from those research libraries whose collections are large and representative of a diversity of disciplines and those who have shown an active interest in or concern for information on developments relating to library automation. Persons were also invited who have had experience in the areas of technology immediately relevant to library automation.

In preparation for the conference a number of state-of-the art papers were written and sent to all of you. These papers may or may not give the answers that librarians are seeking. I fear that those of you who ask "What is my first step?" or "How do I begin?" may not find specific guidelines. The papers, from our point of view, were intended to stimulate thinking and to provoke questions. I hope that you will ask questions of the technologists, and ask them stubbornly, to get the answers you feel are needed.

You, who are the technical experts, will be able to tell us what technology can provide. You will not find in these papers a detailed analysis of the library needs and problems for which remedies are sought through your technology. Therefore, I hope that you will ask questions of the librarians,

[1] See King, Gilbert W. Automation and the Library of Congress. [A report] submitted by Gilbert W. King [and others] Washington, Library of Congress [for sale by the Superintendent of Documents, U.S. Govt. Print. Off.] 1963. 88 p.

and ask them stubbornly, about the requirements of libraries in order to get the answers you feel are needed.

During the next few days we may have occasion to reflect that we, the traditional librarians, may be facing the end of an era. Perhaps the future of library service will have to be entrusted to men who can manage large electronic computers and the mysterious array of machines associated with automation. I do not really fear this prospect, nor do I think that it is entirely realistic. There will be a need to state the intellectual requirements which machines will meet, and I am confident that librarians will develop the insights and abilities to specify these requirements in a way that can be understood by the technologist. I am also convinced that library cooperation, which has accelerated over the past few decades, will be aided by automation.

The Library of Congress has a proud record in providing services which are used by libraries throughout the land and in cooperating with libraries in bibliographical endeavors that characterize the vitality of our contemporary library community. We do not profess to be able to offer answers with respect to automation. Quite the contrary, we are seeking answers and solutions to these very problems. We feel that an investment directed toward fuller understanding of the possible applications from the present and the foreseeable computer technology will be worthwhile. We also believe that the changes that automation may bring will involve the cooperative efforts of the library community of which the Library of Congress is a part and to which it hopes to make effective contributions. It is in this spirit of cooperation that I welcome you to this conference.

Now, to begin the work of the conference, I am pleased to introduce the opening speaker, Don R. Swanson, a noted scientist who has recently been appointed dean of the Graduate Library School of the University of Chicago.

SECTION I

The Library of the Future

Design Requirements for a Future Library

DON R. SWANSON
Graduate Library School, University of Chicago

Information Processing and Libraries

Certainly there are few technological subjects these days that give rise to as many conflicting opinions and forecasts by experts as does library automation. The subject has many facets, and I think that most differences of opinion arise in the choice of a facet.

One of the most provocative aspects concerns the ultra microreduction of recorded information. Presently available commercial techniques permit one to achieve reduction in area of about 500 to 1, but machines which do this barely scratch the surface of what can be done in the laboratory. A good optical microscope can reduce information by a factor of 1 million to 1 in area and there exist experimental recording techniques consistent with that density. With such reduction ratio, a large research library of 5 million volumes could be recorded in the space of 5 books. If your imagination is not staggered by this thought, ponder the fact that electron microscopy in principle permits gaining another factor of 10,000. As physicist Richard P. Feynman pointed out in a talk several years ago at Cal Tech ("There's Room at the Bottom"), one might someday exploit the density of recording information at the molecular level and achieve reductions in area of 10 billion to 1. This would permit putting, loosely speaking, a thousand books on the head of a pin, or all of the recorded knowledge in the world on one or two sheets of paper.

Now let us consider a quite different aspect of automation, the application of computers to information processing. Microstorage in itself is not concerned with the machine processing of information but with the miniaturization of graphic records for subsequent human consumption. Computing equipment for libraries on the other hand is concerned not with the miniaturization of information storage but with processing the data needed for control over and access to information. The distinction is an obvious one, but its implications are important. (This is especially so in evaluating those systems which combine machine-readable search codes with miniaturized graphic storage on a single record.) A rather pessimistic picture has been painted by some responsible and competent scientists who point out that, by and large, computers are less well adapted to library information processing than they are to almost anything else, and that high cost puts automation out of reach for a long time. See, for example, the paper by J. R. Pierce of Bell Laboratories given at the John Crerar Library dedication and reprinted in a brochure entitled *Dedication of the New Building—April 3, 1963* (Chicago, 1963).

With these two extremes of perspective on automation it is understandable that a rather bewildering array of questions arise. Are the microrevolutionists visionary or are their critics lacking in imagination? Are books and libraries here to stay or will we have one day the world's knowlege at our fingertips, immediately and instantly accessible by a few electronic gadgets? Are librarians justified in taking a step-by-step approach, mechanizing specific operations within the library, or should they begin with a complete "system study"? What equipment, if any, should they know about before beginning, or is it all too expensive to justify further attention?

I am not suggesting that these questions are answerable by either librarians or engineers within the time limits of our conference. We might, by dint of sufficient effort, succeed in distributing the confusion more evenly between the two groups.

But to do so we ought at least to begin with some kind of unifying principle or approach.

This approach should be oriented to the user of the library, and we must ask not what we can do in principle but what is needed. For example, instead of reflecting on how much information we *can* store on the head of a pin, we might do better to ask: How much do we *gain* by achieving a 10 billion to 1 reduction ratio instead of 500 to 1? The 500 to 1 reduction ratio has been possible for many years and it permits recording the entire contents of a large library in a few hundred square feet of office space. So, in principle, the capability to dispense with library buildings and to disseminate library materials more widely has been with us for a long time but hasn't yet come about in practice.

It should thus be obvious that the answer to whether or not we shall have "pinhead" libraries within the next few decades does not depend only on what we can achieve in the laboratories. A great variety of complex engineering and economic problems are involved but they cannot be reckoned with until the system itself is designed. In my opinion, the critical questions with which we must begin are these: What would we do with the world's knowledge even if it were at our fingertips? How would we gain access to it and interact with it? What do we want to do that we can't do now?

The Systems Approach

Most of the emphasis to date in library mechanization has been on the application of present technology to traditional practices within libraries. There has been a conspicuous failure to understand the distinction between requirements and design, a distinction fundamental to sound systems analysis. A system may be defined loosely as a collection of people and machines organized for a purpose. The point of beginning for systems analysis is therefore a clear formulation of purposes or requirements *independent* of any particular design for implementation. It is necessary to consider with care what services the library of the future ought to perform rather than to take for granted that they should continue to provide the same basic services only, perhaps, more rapidly and efficiently. Proper formulation of requirements is a most neglected aspect of library systems analysis and this task falls squarely between the librarian and the engineer. Librarians, by and large, cannot be expected to do an adequate job without a good appreciation of what technology can reasonably be expected to provide. Too few engineers take the trouble to look beyond hardware and book-charging systems in order to understand the profound conceptual problems of libraries.

My purpose then is to describe requirements in the light of the possibilities offered by automation and without being constrained by the limits of present technology or by tradition for its own sake. This discussion then is to be regarded as the point of beginning for systems design and analysis and not as a recipe to be taken literally or in detail.

I am going to outline the performance characteristics that we would like to have in an automated library at some future date. Automation is not an end in itself, but to assume that the library is automated forces a more thorough and precise description than would be the case for a nonautomated system. Any lack of understanding of either requirements or performance is readily exposed, and those operations for which human judgment is required can be clearly identified. Furthermore, more realistic estimates of workload and response times can be made than in nonautomated systems; if for no other reason, computers are not subject to Parkinson's Law—a fact of key importance, incidentally, in any enumeration of basic differences between men and machines.

Let us consider the user of a future library confronted with an input-output console which puts him into communication with an automated catalog and other bibliographic tools of a large library or system of libraries. I will assume also the existence of an efficient and fast book delivery capability. This system permits a series of rapid and repeated searches; the console at each stage displays to the user the results of his inquiry, with a reaction time much faster than is found in present libraries. Convenience to the user can be served by proper location of the console—which may be remote from the library itself. The information that can be displayed in a sequence of questions and responses can include bibliographies, abstracts, indexes, tables of contents, etc. The console could include (or have adjacent) a microform viewer, so that selected pages of works requested can be examined first before requesting full copy.

The system just described now provides a framework for a summary of future library requirements. These requirements are expressed in terms of desired performance characteristics from the point of view of the user; thus they represent proposed goals.

Summary of Eleven Performance Goals

1. *User Dialogues; Programmed Interrogation.* Users of present research libraries are largely ignorant of bibliographic tools and information resources. The system described above should operate in such a way as to assist its users to become increasingly proficient as they gain experience in use of the system. Any rational question addressed to the system should evoke a response which instructs the requester as to the type of question he should ask next and which presents him with a set of choices from which he makes a selection. This operation can be described as "programmed interrogation." Successive questions should, by means of such a dialogue, bring the requester increasingly nearer to fulfilling his information requirement even though he might begin with partial information that would be inadequate in a conventional system.

2. *Aids to Browsing.* The shelf organization of books is considered important for browsing purposes in many present libraries, though the notion of browsing itself seems to be altogether vague. Everyone agrees that it exists, but few are able to say what it is. I think it strongly resembles shopping in a supermarket. Without being sure of what it is that we want, the display of wares helps us formulate our requirements and make a selection. Browsing takes place at "walking" speeds and not necessarily at the speeds at which we would like to be presented with information in order to make a selection, i.e. at "thinking" speeds. Furthermore, the real information requirement itself is seldom stated, and we are generally left, if we like browsing, being happy with what it is that we find rather than with what necessarily best serves our purposes. (It is somewhat like philosopher Abe Kaplan's definition of a pragmatist, "If you can't be with the girl you love, you love the one you're with.")

The capability for a rapid response dialogue between user and system opens the possibility of following complex chains of association and, in effect, of performing almost any aspect of the function of browsing that can presently be identified. It is doubtful that browsing in the stacks serves any function that could not better be performed at a console such as is under discussion here. If we examine the type of information that is gained by direct access to the stacks, we may begin to gain some insight as to the nature of browsing. The examination of title pages, tables of contents, and certain other portions of books taken from the shelves represents processes which, except for their physical environment, may not only be duplicated but performed more flexibly and rapidly at a console. Browsing with the help of unusual information clues, such as "chains" of related subjects, use-history, citation patterns, and other means, can be made feasible through well-designed automation. We may note that the need for browsing is one of the principal arguments often given for maintaining a collection of books whose shelf arrangement is organized by subject rather than by some alternative criterion that might permit more efficient storage, retrieval, and delivery.

3. *A User-Indexed Library.* Much valuable information other than subject and descriptive (particularly that which deals with use-history, citations, and user indexing) is not taken into account at all in the design of present indexing, cataloging, and classification systems. Maximal effectiveness of the use of any collection can be achieved in principle by superposing as many viewpoints of organization as is practical. Particularly important are the viewpoints of the users of the library. To incorporate user viewpoints is of course difficult, but with the proper automatic aids it should be possible to do so in at least two ways. First, each time materials are returned to the library the user should also return annotations which reflect his views on the cataloging of those materials, their relationship to each other (the fact that he considers two works as similar in some particular sense may be of great future retrieval value), and their relationship to other biblio-

graphic tools. These should be reviewed by a librarian, be processed in some way, and, if appropriate, be incorporated in the catalog. A second scheme would not require voluntary user cooperation and could be fully automated. Systematic records of who uses which books, periodicals, or journal articles could be kept (at least for those users who grant permission to do so). Any user could then readily recover anything which he (or any other person he names) has used previously. One could ask for "the red book which I checked out last week" or perhaps "the set of 15 books on automation which I used last fall." An entire "private demand library" could be rapidly constructible.

One could also exploit the information that two works which share a relatively large common group of users are conceptually related for that reason. The extent to which such relationships between users and information are of value remains to be seen. By keeping track of the associations between people and recorded knowledge, the library could and should develop a capability to assist a requester in identifying persons who might be able to answer any questions for which the available recorded information is inadequate. Bibliographies could be supplemented by lists of persons who have requested similar bibliographies, and who therefore might assist the user in extending or evaluating a particular bibliography.

In addition, one could imagine automatic "discovery" of new subjects or disciplines through user-book associations. For example, prior to published studies by engineers on "bionics" or on artificial neural nets, it is conceivable that works on neurology and on communication theory shared a significantly large group of users, and knowledge of this would have been valuable in the reorganizing and augmenting of bibliographic tools.

4. *Access in Depth to Information.* Intellectual access to detailed information on a subject basis is, at present, primitive in any area not covered by one of the good indexing and abstracting services. As a rule, subject cataloging of books is little more than a gesture, since a book belonging to 100 categories might well be cataloged under only 1 or 2. Indexing in depth is expensive and the benefits derivable therefrom are difficult to measure. Control in depth to the limits of feasibility should be undertaken in future libraries on as large a part of the total collection as reasonable cost permits. Information centers of specialized types should be set up and maximal advantage taken of techniques for automatic indexing, abstracting, and dissemination. Within certain limits, these have already been demonstrated as useful and feasible even though, for the present, the matter of economics is still open to question.

5. *Wheat and Chaff Identification.* Within our present system of communicating recorded knowledge, no formal mechanism exists for distinguishing between that which is important and that which is unimportant. The difficulties of doing so notwithstanding, it is worth attempting. In future libraries all feasible measures should be taken to identify the more significant publications. These materials might then receive the most thorough indexing and cataloging, to the extent of identifying detailed information content. Available resources for indexing and cataloging are always limited, and within those limits such resources ought to be optimally allocated. Optimization could be taken to mean that the most important material, assuming it can be identified, should be the most accessible to use. Recognition of importance is assuredly difficult, but even a rough or approximate measure may be greatly rewarding. One feasible preliminary step would be to facilitate delivery of reviews or critical comments on published works at the same time the work itself is delivered to a requester.

6. *A National "Network" of Librarians.* We may presume throughout this discussion that limits beyond which machines cannot go will often be encountered. Effective use of librarians is fully as important, if not more so, as effective use of machines. Many reference questions require the assistance of librarians expert in a subject specialty. Proficiency in a speciality well beyond that which most librarians can reasonably be expected to have is often needed, except in relatively rare instances where a librarian may have unusual depth of knowledge in a very few specialties. No one research library in that case could afford specialists in all subject areas. Thus the librarian resources of all major research libraries should in some way be pooled. This could be brought about by means of an interlibrary communication sys-

tem, so that, in effect, any librarian is accessible to any user. The terminal of this communication system should be an intercom or telephone at the user console; this puts the user into contact with the local librarian, who may then utilize a teletype network for communicating with librarians in other parts of the country.

7. *A National Network of Bibliographic Tools.* There has been some recognition that the research libraries and information centers of the country constitute a national network but existing tools (such as the National Union Catalog) and cooperative arrangements fall far short of what could be achieved if a major effort in this direction were initiated. The collections themselves for practical reasons must be scattered, but systems planning, specialized reference services (as above in 6), and bibliographic control (in particular, by means of catalogs) can and should be carried out on the basis of a nationwide, or preferably international, approach. Practical limitations on the size of any collection always exist, but bibliographic tools can be much wider in scope. These tools, in a future research library, should represent the holdings of all major research libraries in the country either directly or by means of a communication network.

8. *Instant Information?* Response time for receiving library materials is inconveniently and perhaps intolerably long compared to the speed of human reactions and requirements. (This requirement is of course difficult to measure since, strictly speaking, *intolerable* response time would lead to a decision *not* to use the library.) The system should, in any event, respond as rapidly to a request *as the requester specifies*, within constraints of practicality. It need not respond with uniform alacrity to all requests since there are those cases for which a response in an hour or two is as good as in a few seconds; there is considerable economic advantage in specifying that the user be given the option of making a response time specification for each request since the cost of providing ultrarapid response all the time is likely to be exceedingly high.

9. *Remote Interrogation and Delivery.* In further pursuit of an "ideal" system we may remove constraints as to the time and place at which a request is made. Remote interrogation and delivery of library materials is of course feasible (even by simple means such as a telephone and messenger service) and is largely a matter of weighing value against cost. The value is intangible, but experimentation may lead at least to reasonable estimates.

10. *Active Dissemination to Supplement Passive Search.* Libraries at present are largely passive and should probably play a more active role in bringing to the desks of users materials which they ought to see. Reading habits of users may tend to be dominated by what is accessible with the least effort, and this argues for maximizing the value of such material. Experimental evidence exists to show that the automation of highly selective, accurate dissemination of at least certain types of material is feasible.

11. *Quality Control Over Library Services; Improvement Feedback.* Finally, and most importantly, there are few systematic procedures built into present research libraries to determine their own effectiveness, i.e. that would permit them to be compared with what they ought to become. Complete satisfaction with present research library services can only arise from a gross misconception of the true state of affairs. Well-designed sampling techniques would make possible detailed information control and evaluation for a small but significant part of the library; the results obtained should then form the basis for improvements of the entire system.

These 11 points form the basis for an approach to the system design of future libraries. Each must be expanded in greater detail; it is beyond the scope of this talk to do so. The user dialogues will be discussed at greater length, however, since these are of particular importance.

Main Elements of a Mechanized Library

Having summarized the performance characteristics required of a future library, let us now consider the main elements of a system designed to fulfill these requirements. These elements are outlined in figure 1. (The performance characteristics will be developed in greater detail as the operation of the system is described.)

The dashed lines in figure 1 represent the flow of digital information in an automated system. All heavy bordered boxes denote automated functions. The flow of books, periodicals, reports, microfilm copy, and other graphical, rather than digital, data is represented by the solid lines between boxes. The user console consists principally of a keyboard, a cathode ray tube (hereafter abbreviated to CRT), which is in direct communication with the computer system and which has a provision for making a permanent copy of displayed information. The CRT displays only digital data stored in the automated portion of the system. In addition, the console includes a microfilm viewer, with controls for search and selection, an enlarger-printer, working space, book shelves, and a small copying device to permit the making of permanent records of any portions of books. An intercom or telephone is also available for consultation with a reference librarian as necessary. These components represent a composite idealized console; the actual consoles should exist in several versions, some very simple and inexpensive, since cost and use factors on different components may differ widely. The console described here is by no means fanciful or beyond the present state of the art; although expensive, such consoles exist now. For the most part they have been designed for specific military applications, but their characteristics are similar to those required for library use.

The box labeled "Union Catalog" represents, in principle, a universal catalog of knowledge. In practice, however, it might reasonably represent the holdings of all major research libraries in the country. (It is hoped that, eventually, a national union catalog will be available in machine-language form, periodically updated and maintained at the Library of Congress, and made accessible to other research libraries.) The union catalog proposed here, however, will differ considerably from the present NUC in accordance with the requirements outlined earlier. In particular, material of high quality will be cataloged in greater depth than the rest, and the catalog will contain the ac-

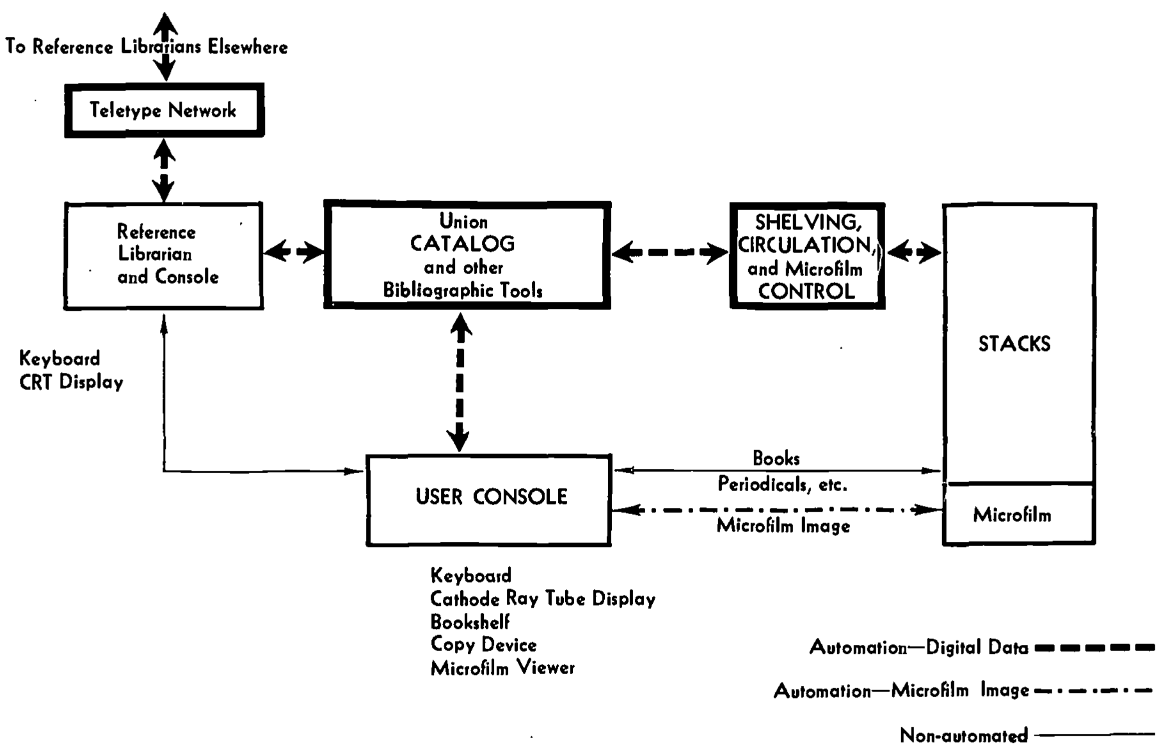


FIGURE 1. *Main elements of an automated library system.*

cumulated information put back into the system by means of annotations supplied by the users of library materials. "Cataloging in depth" implies that all known tools of descriptive and subject access will be available for experimental comparison and for testing of retrieval efficiency.

The box labeled "Shelving, Circulation, and Microfilm Control" is of particular interest in that its lines of communication run to the catalog and to the stacks and consist entirely of digital data. Thus, shelving and circulation control is achieved automatically insofar as the signals sent between stacks and catalog are concerned. Similar signals must flow between user and catalog to complete the control function. Physical retrieval of books, documents, and periodicals may or may not be automated since this is rather difficult to implement and no satisfactory scheme for automatic book retrieval has yet been proposed. Our system here, however, does presume that the *control* of books is automated, although their physical handling may not be.

The reference librarian shown in figure 1 is also provided with a console. This console consists of a keyboard, CRT display, a voice intercom to user-consoles, and a teletype network to other librarians in the country—in particular to those best able to provide expert reference assistance in specific disciplines where more detailed knowledge is required than the available reference librarian may have.

The microfilm library noted in figure 1 is to be regarded here as a step intermediate between the card catalog and the complete work itself. One may reasonably assume that in many instances the user cannot ascertain from the catalog card whether he actually wants to see the corresponding book or periodical. It is reasonable to think that he could avoid calling for the book if he had access (in microfilm) to its title page, table of contents, index, and perhaps a few selected pages. Complete wo' in microfilm may of course be stored to the extent economically feasible and acceptable to users.

To develop the proposed system further a detailed description must be given of all operations to be initiated at the user and librarian consoles. The amount of detail needed is enormous, but it will be useful to begin with a narrative description of the interchange between the user and the console-computer system. Before such a system could be implemented, the computer to which the console is connected would have to be programmed to respond appropriately to each signal sent from the console. It should be recognized that the following description of processes might well imply 50 man-years of programming before the system could be operational.

Main Control Keys for Initiating Dialogues

The following description of some of the basic operations from which dialogues between the user and the console can readily be constructed should not be considered definitive; it is intended to exemplify the kind of interchange that might reasonably take place, and which clearly holds the potential for a much more penetrating interaction with the library than conventional systems permit. The console instructs the user on a step-by-step basis on how to proceed in his task of interrogating the library in order to find information which he requires. Each question is asked of the system only after some kind of response to the previous question has been given. "Programmed interrogation" is an appropriate description.

Six major "process control" keys present the user with his initial set of choices at the console. These keys (illustrated in fig. 2) serve to specify the major type of operation to be performed.

The key labeled "Specific Work" is to be used to specify a particular work by the usual (and some unusual) descriptive material, such as author, title, publisher, date of publication, etc. The requester need have only partial bibliographic information at hand in order to obtain a response from the system, and the system will present him with a bibliography of all those works which meet the criteria that he specifies.

The key labeled "Subject Selection" permits retrieving material based on a particular subject, retrieving material responsive to a specific question, or browsing. Browsing is implemented by means of the subject key, the combination key, and the similarity key. The combination key is of particular importance since it permits successive operations with several keys in order to combine a set of retrieval specifications.

The "Previous Use" key permits the requester to recall any books that he has used before, or that some other person he names has used, even though

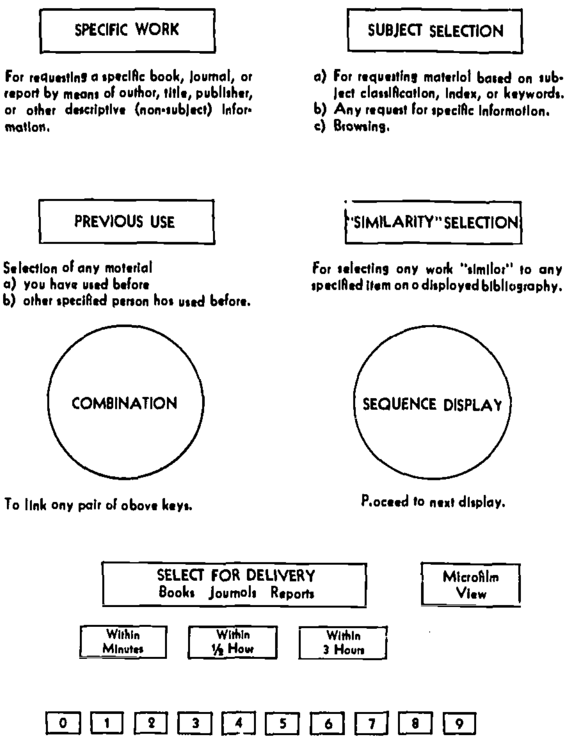


FIGURE 2. *Main control keys at user console.*

he cannot remember enough conventional descriptive data to specify them uniquely.

The key labeled "Similarity Selection" initiates a chain of bibliographic citations that satisfy certain criteria of similarity to any initially specified work.

The "Microfilm View" key is used to call for a microfilm display of selected portions of any work identified on the CRT display.

In designing a computer system to respond to the console, the time delay permitted in any response is a very important design parameter; if overspecified it is possible to run up automation costs by several orders of magnitude. The cost of providing a 1-second response to every signal might well be 10 times greater than the cost of a 1 to 10 minute response time. As noted in our list of requirements, response time clearly should not be specified as a fixed amount, but rather the user should have the option at the console of selecting what he needs from a number of possible choices. An example of such option is indicated in figure 2 by the three keys below the "Select for Delivery" key.

In the following paragraphs we shall trace the consequences of activating two of these major control keys and show the conceivable chain of questions and responses which could result. The immediate result of actuating any particular key is to receive a display on the CRT which presents the requester with both information and instructions on what to do next. The end product of each sequence is a bibliography from which the user then makes a final selection. Prior to receiving the bibliography he is generally presented with information on how many items the bibliography will contain. He may then narrow the search by further specification, if necessary, to prevent receiving an unmanageably long list.

Request for a Specific Work; Tolerance of Ambiguity

Possibly the simplest and most frequent type of library request is that for a particular book or work, identified by title and/or author and/or other descriptive bibliographic information. The system should be designed to operate on ambiguous, incomplete, and/or unconventional bibliographic information. No matter what is specified, the system should deliver to the requester a list of those items which fulfill the specifications and from which he then may make a selection. If, for example, the requester can specify the last name of the author, but nothing else, it should be possible for him to receive from the library a complete list of the works of all authors with the specified last name. If he specifies only the title he should receive a list of works having that, or a similar, title. Other information, such as publisher, date of publication, identification as serial or monograph, etc., must also be acceptable to the system in incomplete form. The system should reply first by stating the *number* of bibliographic citations to be listed, and then by producing a bibliography if one is desired by the requester. Each bibliographic entry should identify alternate sources for the work requested if the work is not in the collection at hand or is otherwise unavailable.

If the requester is uncertain as to the suitability of any particular bibliographic item, and if that uncertainty could be removed by examining the table of contents, the index, or perhaps the first page of the work, then he may ask for such pages in microfilm copy. This he does by means of the "Microfilm View" key. Finally, the "Select for Delivery" key is used for delivery of the entire work—either in book form or microform. If a requested work is not available for delivery, the user is given immediate and complete information as to its status (i.e. when it may be available, whether recallable, whether in use at a nearby console, etc.).

Response to this type of request (illustrated in fig. 2) can, in principle, be fully automated, even with only a fragmentary bibliographic citation unless the requester has incorrect or badly garbled information. In that event a librarian may be able to help him reformulate his request. In general, however, the automated system is designed to tolerate as much ambiguity, vagueness, and lack of information on the part of the requester as possible.

Subject-Oriented Requests

Requests for bibliographies of works on a given subject are considerably more complex than those based solely on descriptive information. In this case the problem is not necessarily susceptible to the same degree of formal description, and a conference with a librarian may be necessary. Suppose we try, however, to analyze this requirement into its basic intellectual components insofar as possible.

The requester begins with some kind of specification formulated into words without regard for any preestablished subject categories. The process of transforming the initial requirement into a list of possible subject headings or categories is an intellectual task involving concept association. A good part of this association, however, is achievable by means of word association in a rather mechanical fashion. Conventional alphabetized subject heading lists serve this purpose through "see" and "see also" references. In such a dictionary list (used, for example, in the Library of Congress catalog) the requester looks up a word descriptive of his requirement and finds in general a "see" or "see also" reference to some subject heading under which he can then find a bibliography. Word associations by this means (or by means of a thesaurus) are more successfully carried out if the scope of the subject matter can initially be narrowed. Therefore, in the machine dialogue we will assume that the requester is first presented with a gross or major classification of human knowledge from which he may select the most appropriate categories. Following the selection of the major subject, he then enters a string of keywords. Since some of the words will be more important for purposes of the stated requirement than will others, these words should be entered in their order of importance. The machine search procedure will then take advantage of this order of importance in arranging the sequence of the bibliographic listing. The method of constructing such a sequence permits the well-known problem of forming "logical combinations" of terms to be circumvented.

With this string of keywords and a selected discipline, a machine matching procedure will (before constructing a bibliography) construct a list of subject heading groups from which the requester may then make a selection of those he thinks appropriate.

Before calling for a bibliography based on the subject headings and keywords specified, the requester should be presented with a list of the number of bibliographic items in various categories of his search request, determined by the number of specified terms which are satisfied. That is, he should first be told how many items match all of his specified terms (here a term means either a subject heading or a keyword) and then the number of works in all but one of his specified terms, all but two, etc. Thus the requester is not burdened with the exceedingly difficult task of making a judicious choice of "logical combinations" of search terms, since he is presented immediately with the consequences of all possible such combinations. From these numbers the requester can judge whether or not a further narrowing of the request is necessary. If it is, and if he has any further descriptive information or can supply information as to previous use, he may initiate these specifications by first activating the "Combination" key and then entering the appropriate new sequence. When the bibliography length is finally acceptable, the requester calls for a bibliography and then selects either microfilm portions or complete works in the same way described for requesting a specific book.

A system which responds to the user with a bibliography at each stage of his request serves a valuable function as a training device in the use of the library. The requester not only learns that a poorly formulated request does not lead to responsive information, but he also learns a good deal about what was wrong with the request, e.g. whether it was too broad, too narrow, or too ambiguous.

The Economics of Automation

One need have only slight acquaintance with computers to recognize that on-line rapid response access to a large catalog, as is implied by the discussion in this paper, is enormously expensive, if not prohibitively so, with equipment presently on the market. It is, however, the prerogative of the customer to specify what is needed, and it would be an abdication of this prerogative to abandon desired requirements, before they are formulated, on the grounds that they might be too expensive to implement. Engineers should first be given the opportunity to design systems that meet the requirements and for which a competitive price tag can be established. Judgment can then be exercised as to whether the probable benefits are worth the price.

The hazards of limiting one's thoughts to current technology can be illustrated by the following example. Approximately 2,000 reels of high-density magnetic tape would be needed to store 10^{11} bits of information, which roughly represent the size required for a future National Union Catalog in the Library of Congress. If each of these reels were mounted on a tape unit, so that any part would be accessible within 4 minutes, then, at $20,000 per tape unit, the cost would be about $40 million. At least one manufacturer now claims a capability of developing a large capacity memory with the following specifications compared to the above tape system:

1. A storage capacity 10 times greater than the the capacity of the tape system.
2. A cost of only one-tenth that of the tape system.
3. Provision of access to any piece of information in a fraction of a second, rather than in the 4 or 5 minutes required by the tape system.

It cannot be determined at the present time whether or not this claim is accurate, but there is at least good reason to say it is not irresponsible. While it may be argued that no one has seriously proposed putting the catalogs of the Library of Congress on magnetic tape, it is likely that the capabilities of tape equipment have influenced much of the current thinking about the inappropriateness of automation for library operations.

In some respects the same potential for future improvement may apply to console equipment, which at present is quite expensive. Highly versatile consoles have been designed and produced (in quantities of 1 or 2) for prices as high as $200,000 each. If procured in lots of 100 or more these same consoles, it is estimated, would cost perhaps $30,000 each. New approaches to information display have been suggested which might conceivably permit eliminating a large portion of the electronics in a console, i.e. that part associated with the buffer required to refresh the cathode-ray-tube display. If this were done, costs might be radically reduced. Whether or not any company decides to make such a console may depend on whether the requirement (hence the market) is important.

The point of this discussion, again, is not to contend that economically feasible automation is just around the corner; rather it is to emphasize that the customers of library automation must formulate requirements and thus cause manufacturers to find economical solutions. There may be room for improving the present economic picture of automation by a factor of 10 or more. One gets the impression that some of those who are designing automated libraries that cannot possibly be implemented in less than 3 to 6 years are basing their requirements on the performance of today's libraries and their design on yesterday's computing equipment.

Automation and the Library Profession

I would like to close with a note on automation with respect to the profession of librarianship. In the field of librarianship two pervasive questions exist that are closely linked, although this relationship does not seem to be very widely recognized. The first, which has been with us for many decades, is concerned with the distinction between the professional and the clerical aspects of librarianship. In what sense is librarianship a profession and in what sense does one need training only in certain routine clerical practices? The second question—is the automation of libraries possible?—is, of course, the one that we are dealing with here at this conference. I think there is one answer common to the two questions. Those library operations that are reducible to clerical routines are those which are mechanizable. This sounds obvious, but the trick, of course, is to recognize and identify those which really are reducible to clerical operations. Having done this, one is still left with tasks which require a high level of professional judgment.

In my view, therefore, automation is far more likely to upgrade the profession of librarianship than to replace it. Automation upgrades it by permitting a sharper and clearer identification of that which is really of professional character in librarianship. Those librarians who have some kind of irrational antipathy toward mechanization *per se* (not just toward some engineers who have inappropriately oversold mechanization) I regard with some suspicion, because I think they do not have sufficient respect for their profession. They may be afraid that librarianship is going to be exposed as being intellectually vacuous, which I don't think is so. Even in a completely mechanized library there would still be need for skilled reference librarians, bibliographers, catalogers, acquisitions specialists, administrators, and others. Those librarians in the future who regard mechanization, not with suspicion, but as a subject to be mastered will be those who will plan our future libraries and who will plan the things that machines are going to do. There will be no doubt of their professional status.

735-898 O—64——3

CONFERENCE SESSION I

General Discussion

FUSSLER: I am interested in your distinction between that which is professional and that which is clerical. I understood you to say that that which is mechanizable is clerical and that which is not is the realm of the professional. I can visualize, for example, a computer program which would diagnose a disease which is normally a doctor's professional responsibility. Does this mean then that you would be willing to delegate to the medical profession a clerical activity?

SWANSON: Yes, as a matter of fact if medical diagnosis is mechanizable, it means that doctors were performing in a machine-like manner, without, perhaps, being conscious of this. If you can feed in a string of symptoms which invariably lead to a diagnosis, then this can be taught to a machine. It may have required professional judgment to make this particular discovery, but, having made it, it then becomes a machine-like task, clerical if you will, to apply or implement it.

I should note, by the way, that there are some hazards in equating "machine-like" with "clerical." All human beings, including clerks, have, for example, a highly developed facility for complex pattern recognition which is used in descriptive cataloging, in handling books, and in other library operations. This certainly represents an activity, at the clerical level, that technology isn't yet able to mechanize.

FUSSLER: One more question. The original decision, or evaluation, may be considered professional, but subsequent repetitions of that decision or evaluation, therefore, become clerical?

SWANSON: In principle, yes. Now I can see a qualification I should add. I suppose one can arrive at a clerical decision by an intellectual process. There may well have been doctors going through such a process to produce the end product in our diagnosis example. It may well be that the end product could have been produced by a clerical-like procedure, regardless of how intellectual any doctor might have been in arriving at it.

ALEXANDER: May I add a footnote to that? You do have the problem of assuring yourself that the process is deterministic—that for a given set of inputs there is one and only one repeatable set of correct answers. When that happens, no matter how complex it is, you can argue that you can devise a machine program to reproduce it. The difficulty is that many of these situations are not fully determined. Consider the extreme example of manuals of procedures—legally there is one, and only one, correct answer on how to proceed if you follow the manual. This becomes a very complex, but clerical, operation. Your job is not to make the decision but to find the page on which the decision is recorded. I think that this analogy sometimes gets you out of this impasse. If the activity has not been reduced to a set of procedures so that the response is determined when the input is given, then you don't have a clerical or robotlike response.

The process of writing the manual of procedures is a very highly professional task. The more you put in that manual, the more you have built a repetitive structure that is clerical in character. We usually think of clerical activities in terms of the operations being done, rather than the choices being made. There is a class of activity which purely is: On what page will I find the answer to this set of circumstances?

SWANSON: It should be added that, even though the answer is deterministic, it doesn't necessarily mean that it has to be unambiguous, in the sense that the range of ambiguities can be presented for human inspection at the point of output.

FUSSLER: What you really mean by clerical is that it hasn't been analyzed to the point where a decision can be made and you have actually worked out the decision. For instance, you said that the diagnostic machine is clerical once you've done it, so what you really mean is you've gotten to the point where you've analyzed it.

SWANSON: Yes. And perhaps one can infer from this that the proper professional activity for librarians is planning things for machines to do, because it is the writing of the manual of procedures that requires professional knowledge. Too many librarians are *following* written manuals of procedures and therefore are mixing together the professional and nonprofessional tasks in an indistinguishable way. This has been the concern of the profession, and the failure to make a profession out of it, in many cases.

TAUBE: I think that it's rather unfortunate to use the word "clerical," because within the library profession a clerical job is something we look down upon, as opposed to the professional job. You wouldn't say that mathematics is a clerical job—it is a formal job—and the computer does very well in mathematics. Now I've been known to talk about the limitations of machines. But I would, though, rather than the word "clerical" use the word "formal." Those processes in a library which can be formalized and treated according to rules can be treated with machines. Therefore, it is a major intellectual job of librarians, following what you said, to reduce to *formal* procedures those things which they have been vague about up until this time.

SWANSON: I think that I like that term better than clerical. The reason for the use of clerical was to make clearer my assertion that this is related to the issue that has pervaded the profession of what is professional and what is clerical. The word "formal" has never come into the language in that connection.

BORKO: I wonder if you would say something about indexing and classification in terms of finding relevant material in this library of the future.

SWANSON: I presumed that in the library of the future one superposed as many schemes of organization as are economically practical, that is, indexed, cataloged, or classified to the depth that was economically practical. I suggested that perhaps one allocated limited cataloging resources to the more important material, if that were possible. I suggested incorporating, taking maximum advantage, of the users of the library to help organize the library materials. I purposely did not try to pin it down more closely because this is an enormously complex question. Harold, you're an expert in this area; maybe I'm missing something that you can add.

BORKO: No, I become frightened, even with the nano-second computer, at the thought of going through all of the material in the National Union Catalog at some real-time response. We have to find some way of breaking this down, of indexing our material, so that we can tag it so that the computer can give it to us to browse.

SWANSON: I agree it's frightening, but there is a danger in being frightened when one is trying to state requirements, because it might turn out that we are frightened about the wrong things. After all, it is now a problem for the engineers to solve this business of nano-second access to an infinite amount of information. We put aside, within reasonable limits, the question of the economic practicalities of how much information is to be handled how fast and for what cost, until we have definitely ascertained that we can't have what we want, and then we take Kaplan's view of the pragmatist and are satisfied with what we have.

UNIDENTIFIED SPEAKER: In this future library how many users could be making simultaneous use and how close together would these uses be? In the present sedan-chair approach to the Library of Congress for browsing, there is room for almost everybody. It may be clumsy, but they can get in.

SWANSON: I don't think that this is any more than the conventional waiting line or queuing problem, in which you have a certain number of service points. There should be room for everyone in libraries in the future. Everyone may not be able to sit at a console, and furthermore all these consoles may not look alike, because after you have studied the use patterns you might find that an expensive printer, for example, should be shared among five consoles; the microform viewer with an ultrasimple keyboard device, among two. I suspect that this console gets fragmented into four or five different types of consoles, and that the numbers of each are adjusted to the relative workloads in various parts of the system thus minimizing the idle time of any particular element. When all is said and done, you will still have desks with no consoles at all, because there will be the person who wants to sit and look at the hard copy that came out. He cannot tie up a console while he's doing that, but he does need working space. So I think that this is more a conventional systems study problem, where one makes an optimal choice of the quantities of equipment to fit the workload in the individual parts of the system.

DIX: Dr. Swanson, in this system that you have described, the ultimate product is a piece of prose—am I right? That is, it might be a picture, a table of contents, a piece of music, but basically it's a piece of prose. In other words, it doesn't attempt to go beyond identifying the items requested by the use of symbols relevant to a particular topic. I think some of us have been confused even by the very term "information retrieval" and were thinking of some proposed machines that would answer questions. As I see it, the only question that your machine answers is a biblographic question. In other words, it answers the question: What does this system contain in prose relevant to this topic? The problem of analyzing what is in that piece of prose still remains a human problem. Now if you go beyond this by depth of indexing, it might tell you what page it's on in that piece of prose, or what line it's on, but it will not print out an answer to a question that might be phrased like this: How do I get inkspots out of my tie? It will produce for you the literature written on that subject. Is this right, or am I oversimplifying?

SWANSON: No, I think this is correct. I have talked about a system that essentially provides bibliographic information. It does a few other things, such as suggesting to the user how to ask for references so as to get the bibliography, but the end product is essentially a bibliography.

ALEXANDER: Is it not true that the analysis is the crux of the whole system, and is not the analysis, which is done by humans, the most important element of the whole system?

SWANSON: I didn't mean to imply that this was not so; if I did, it was in error. It is certainly true that the finesse with which one follows intellectual routes to information depends critically on the indexing, cataloging, classification, and organization of the information. The only observation which I think appropriate to make within the framework of this talk is that, in principle, the more viewpoints one can superimpose on a collection, in terms of the way it is indexed, cataloged, organized, and classified, the more effective the retrieval, and the user viewpoints are among the most important to consider. How one goes about superimposing a lot more things than one has the money to superimpose and how to get at the information in a lot more depth than one can afford to are engineering problems. I am inclined to the view that, in principle, we now know how to index and classify about as well as we are ever going to. It is more a matter of engineering and economics to get all of what is known into the system, so that we are not in the position of the farmer resisting new ideas because he "ain't farming half as well as he knows how now."

ALEXANDER: I would certainly agree with that, but the result is that this becomes the most expensive part of your whole operation, much more than a $70,000 console.

SWANSON: I agree. Particularly if one thinks in terms of indexing every sentence in the Library of Congress. The problems are tremendous in reducing this to something that is economically feasible, if you consider everything that could be done in principle.

ALEXANDER: Isn't it a problem of how large the aggregates are? A sort of compromise would appear necessary between the librarian trying to meet the needs as he sees them and the engineer trying to stay within the economic restraints.

SWANSON: What you are saying is that we have to compromise, and that is a matter of ingenuity, as well as a willingness to compromise the requirements. But isn't this really the heart of systems analysis?

ALEXANDER: We must recognize that the search strategy you use with a combination of machines and people is not necessarily the search strategy you use with people alone. Our difficulty really is that we are trying to use search strategies we inherited, when we need people who have learned the delicate art of mixing the two types of search strategies.

SWANSON: What I hope we can design (and I'm not pretending that I *have* the design) with this particular approach is a system which makes it easy *not* to be bound by tradition in the sense of giving the user sufficiently flexible tools to explore and discover search strategies that are not traditional. Do you have any specific suggestions as to how we can go further in that direction? Do you think the system as described here is, perhaps, overly constrained toward causing the user to use strategies that he has traditionally pursued in the past?

ALEXANDER: I'm afraid I have raised a question rather than answered one. But I would like to point out that, unless we follow different trends then we have, we will start by building in the search strategies with which we are familiar, and then we may find out that they are either too expensive or too clumsy. This certainly has happened in trying to apply the same machinery to administrative procedures. The buggy whips went in on the automobile right from the start and it took quite a while to get them off.

BUCKLAND: Aren't you providing more modes of access to the library by shifting some of the cataloger's knowledge and some of the librarian's judgment over to the user?

In many ways your query-and-response console system is an area akin to teaching machines. Here two things have emerged. One is that paper machines are just about as good as the ones involving time-shared consoles, and the other is that they are successful when they cover areas of information that are extremely well mapped and have been well analyzed by lots of people.

SWANSON: I didn't mean to suggest that one shifts the burden from the librarian to the user; however, I see no reason why we should not utilize the contributions of both as far as it is feasible in the system. Of course, someone has to catalog, classify, and organize the library. There is adequate room in this system for making the most effective use that we can of the librarian and his resources and then simply superimposing whatever the user can offer.

SECTION II

File Organization and Conversion

Index Files: Their Loading and Organization for Use

ROBERT L. PATRICK, DONALD V. BLACK
Planning Research Corp.

Introduction

A well-intentioned but ill-informed salesman or other "expert" may someday attempt to convince you that libraries should be mechanized by librarians. He may quote from the manual on latest "magic language" and point out that anyone can program for his machine using the latest programming aids. On the other hand, some overzealous practitioner of the "black art" of computer programming may take the egotistic view that he knows enough about libraries to mechanize a library without any help from the staff. Based on experience, the probability of either of these being correct is quite low. At the present time both fields are so complex and changing so rapidly that no man can know all of either, much less both. In recognition of these facts, the coauthors bring to bear the combined skills of the library scientist and the computer specialist. This interaction has proved so valuable that the paper itself was framed to allow our readers to begin establishing a similar relationship with a qualified individual in their own locale.

A fairly thorough literature search brought to light many references concerning the automation of libraries. (See items 3 and 4.)[2] In digesting the documents found, two observations became evident. First, the computer field is quite young, suffers from the lack of an accepted glossary (see item 2), and lacks any codified set of principles on which decisions may be based. On the other hand, the library field is well established, has relatively well-defined problems, and is proceeding (at some undisclosed pace) toward their eventual solution. In all of the retrieved literature, not one document was found that discussed the establishment of the basic files for retrieval. All authors assumed these files into existence. Therefore, it is the neglected topic of basic files to which this paper is addressed.

Assumptions

Before we proceed, it behooves us to narrow the field somewhat further. Of the problems that beset libraries, one of the most pressing seems to be money. This, and its ramifications, will be discussed in Dr. King's paper. Therefore, it is assumed that an automation program has been justified and we will proceed to indicate how this may be carried out.

In addition, there are several little technical problems that we, like other authors discussing library mechanization, assume out of the way. While we realize that basic technical difficulties do exist in the areas of cataloging and indexing, and that sincere researchers are attacking them, we must pass over these difficulties as being beyond the scope of this paper. Therefore, we will assume that an acceptable scheme does exist for the subject analysis of library materials. The scheme may be implemented, as it is now, by human beings, or it may be computer based. In either event, we assume that the document is indexed[3] according to some classification and/or subject heading scheme, and that the depth of such indexing is as great as resources allow.

An additional assumption was made concerning the hardware to be used for the storage and manipulation of our file. Although there are some exotic devices in the developmental laboratories, some of which even show real promise, these have been ignored. We have limited ourselves to pres-

[2] This and similar references refer to items in the bibliography, page 48.

[3] We will use the term indexing to mean the practice of cataloging and classification as normally carried on in libraries, especially subject cataloging and the assignment of subject headings.

ently available commercial devices; however, we have allowed ourselves a reasonable expectation of the future state of development of these devices as their evolution continues over the next few years. Thus, we are considering computers of the present transistorized, solid-state variety. We presume present technique in our discussions of punched-card reading, magnetic-tape transports, and disk-storage devices. There we stop. We have not presumed any exotic computer organizations or breakthroughs of any sort. Our assumptions are all within an order of magnitude of commercial computer components that are presently available "off the shelf." As a further matter of practical necessity we discuss specific, competitive, generally available items. Our choice of items should not be considered an endorsement of a particular manufacturer or as a recommendation for a specific device. The ones chosen are devices that the typical librarian might be expected to encounter in the normal pursuit of his business.

Definitions

Libraries.—A library has been defined as a collection of books kept for use, study, or reading. Some people find it advantageous to differentiate between special libraries and general reference libraries. Others find it beneficial to differentiate between serials, monographs, technical reports, documents, books, engineering drawings, sheet music, stereo records, audiomagnetic tapes, patent disclosures, or letters of correspondence.

While these various distinctions and classifications are of use to the administrator and the librarian, the computer programmer prefers to think in terms of a more general definition: a library is a collection of information. To be sure, the other designations have utility and meaning, but the factors of more interest to the programmer are the size of the collection as measured by the total number of items in the inventory, the kinds of transactions that relate to the items in the inventory, and the time series of transactions over some statistically significant measurement period. Such a time series encompasses, of course, the frequency of each transaction type, the total number of occurrences by transaction, an analysis of the periods of dwell, and the peakload phenomena.

Another useful distinction is the difference between a library and a warehouse. A warehouse is an establishment for the storage and protection of items until called for. The calls for such items are singular, unique, and virtually preordained at the time of initial entry into the collection. On the other hand, a library is an indexed collection of information. One or more indexes to the collection may be developed and maintained. A call for an item may not be unique or unambiguous and seldom is singular.

Index Files.—From the distinctions given above, we may define some useful terms. If the complete description of an item of information is known and if its whereabouts is to be determined, an index file may be referenced by an operation known as a "selection." From this we may imply some things about the specification of an index file. An index file contains a description of the item and its probable location. In consideration of selection efficiency, an index file is usually established, kept, and maintained in some order.

In the case of a warehouse the index to the contents of the warehouse is maintained in the same order as the warehouse receipt (which is approximately date and time). In the case of a library the index file is usually maintained in "alphabetical" order ("alphabetical" here is defined as being suitable to human beings and not exactly in strict alphabetical sequence as defined by a computer sort on a specific unique character set). For a library three such indexes are often kept: an author file,[4] a subject file, and an inventory or shelflist.

You will note that index files can be discussed without engaging in the debate about the media on which the collection itself is stored. The discussion of hard copy or microform is the province of Dr. Alexander's paper. Our previous restriction to state-of-the-art techniques further precludes detailed discussions of natural-language text held in magnetic form. We limit our discussion to the index of a collection of information. The index consists of the item descriptions and the probable location of each item.

Lest the scope of our discussion appear too narrow, we might pause to mention some specific applications. The first one that comes to mind is, of course, the Defense Documentation Center (for-

[4] We here define the author file as being a "main entry" file.

merly ASTIA). The Bureau of Ships has a similar problem with its technical reports. Many military agencies have the same problem in the control of their classified reports. The Los Angeles County Sheriff has the identical problem with a file concerning criminals.

In each case an index file that consists of descriptions of items and their probable location is involved. The items may be books, technical reports, or, as in the case of the Los Angeles Sheriff, people. In the latter portion of this paper we have chosen as our example the National Union Catalog. This was done for several reasons. First, it is a very large index. Second, a conversion and handling technique for a large file can be adapted to work for a smaller file, although the converse is not generally true. Third, the National Union Catalog has some interesting side problems which allow us to discuss character sets, encoding and split files, and to provide some basic information for the state-of-the-art paper on output printing. Fourth, this file presents more taxing problems than other files, such as serial records, shelflist, acquistions list, etc.

Fields.—A file is a collection of entries. An entry is a collection of fields. A field is a named item of information. In addition to being named, each field must be defined. It is defined by giving its mode, its length, and, if it is numeric, its scale factor. In defining the mode of a field, a programmer would note whether the field is alphabetic, numeric, or alphanumeric.

If the field is numeric, the programmer must indicate whether the information is binary, 4-bit coded decimal, or 6-bit BCD (binary coded decimal). If extended character sets or alternate character sets are sometimes used, this, too, must be indicated.

To complete the definition of a numeric field, the programmer must indicate whether the data is signed (e.g. plus or minus) or unsigned (e.g. always plus), the length of the field in either bits or characters, and the scale factor, if any, that is inherently associated with the field. Of course, any or all of these factors can be generally provided by convention. Provisions must be made for all of the exceptions expected.

If a field is alpha or alphanumeric, the programmer must define the character set and its encoding. The encoding denotes the sort order, a topic to be discussed in more detail later. If only uppercase letters were required, a full alphabet of 26 letters, 10 numbers, 27 special characters, and a blank character could be encoded in the commonly used scheme for 6-bit BCD. If more characters are required, then a 7- or 8-bit code also could be utilized. The Federal Government, the military services, the manufacturers' society, and the computer community are now engaged in an extended debate on just what the characters in a 6-bit set should be and how these characters should be encoded within the set. There are already almost more variations on this basic set than there are rabbits in the western United States.

Character Sets.—Even though we have not yet agreed on a standard character set for our data processing computers, each computer has a single set built into it. This set is the "natural" set of that computer, and all other sets are defined in relation to this built-in character set. The character set is extremely important in file definition since the character set defines the order of a file once it is sorted.

Most computers sort by a simple comparison of the binary patterns that represent the fields of interest. While this sounds complicated, in reality it is not. For illustration, if 6 binary bits are used to represent a single character, as in most of our present-day computers, the natural character set of the machine will encompass 64 characters. There are as many characters in the character set as there are discrete states in the representation, i.e., $2^6=64$. Thus we find that a 6-bit code will allow 64 characters, a 5-bit code will allow 32 characters, and a 7-bit code will allow 128 characters. There is uniform agreement that one of these characters must be blank—the communications people call this master space. There is general agreement that we need to encode the 10 representations of the Arabic digits 0 through 9. Furthermore, there is reasonably general agreement that the 26 letters of the Roman alphabet should be encoded. Beyond this there is no general agreement. (See item 1.)

Of specific interest to the library community are the encodings of the punctuation marks and special characters. Once these have been chosen, the sorting sequence for the computer is defined. The computer merely compares the binary patterns character by character (or word by word), and, by a long selection of binary choices, orders the file on the fields that it has been programmed to inspect. A computer could order an index file alphabetically on the author's name field and thereby achieve what we usually call an author index (in alphabetical sort by the name of the author). The computer could be instructed to sort on the keywords and an index file by subject would result. Or the computer could be instructed to sort on the call number,[5] and a shelflist would be obtained.

The character set is the foundation for the ordering and sequencing of a file. Fields are defined by mode, length, and scale. An entry can then be defined by indicating the names of the fields contained in that entry. With such a general technique a further breakdown of entries by entry type can be obtained. An entry type is a unique combination of defined fields. If the definition of a field or the collection of related fields is altered, a new entry type must be defined.

Fixed vs. Variable.—Before we proceed with examples using these definitions, three more words are required. The word "fixed" is an adjective in computer parlance used to denote a feature that is, for all intents and purposes, permanently situated as presently described. Some computers are described as fixed word-length computers. This does not mean to imply that their word length could not be changed, but merely that the difficulty associated with changing the word length of the computer would be severe. In a similar manner, the length of the field containing the year of publication for a book would be fixed at four decimal digits. This does not mean to imply that this could not be modified after the 99th century, but that such provisions are ignored for practical reasons.

The word "variable" stands in computer parlance as an antonym to the word "fixed." If an item of information is considered variable, then provision is made so that the current length of the item of information is indicated in the computer memory immediately adjacent to the item of information itself. The variable-word-length computer determines the length of a data field anew each time that field is manipulated. Such an indication may be a special character in the character set used for indicating the end of a data field, or it may be carried in a separate length field, which is stored adjacent to the item of data. In the latter case, the length field itself would be fixed, whereas the data field would vary in response to the quantity stored in the length field.

File Design

Now, let us broaden the scope somewhat. If the statistical characteristics of a file are not known, then the designer of a file must make allowance for the longest occurrence of each data field that is possible. He then defines a preliminary file organization allowing only entries having fixed length fields. Their length is set, in each case, to the maximum length expected. As the file designer becomes acquainted with his file, through direct contact or through sampling by a computer, he performs a statistical analysis to obtain the number of redundant blank characters associated with each data field. He is interested in the actual maximum length that occurs in each data field, the actual minimum length that occurs in each field, and the frequency of occurrence of each length between the maximum and minimum. Once he has access to this information he then can do an intelligent file design.

We have three very powerful adjectives: fixed, specified, and variable. They will be exemplified below, using the length of a data field as an example. This will be a convenient mechanism for examining the three words whose concise definitions follow. Fixed implies an item whose characteristics "never" change. Specified describes an item whose characteristics change infrequently and are arranged in neat sets of like characteristics. Variable describes an item whose characteristics are prone to change with each occurrence.

If the maximum length and the minimum length for a given field are the same, then that field is, naturally, a fixed length field. The fixed length is defined as the observed length. If the maximum length and the minimum length vary appreciably, then the programmer looks to see what power of 2 is required to encompass this variance. For example, if the number of characters varied from 27 to 56, then the variance would be 29 characters. Since 2^5 is 32, five binary bits would be sufficient

[5] Letters, figures, and symbols separate or in combination assigned to a book to indicate its shelf location. It usually consists of a classification number and book number.

to describe this excursion. The programmer would then contemplate adding a 5-bit fixed field to the data field in question and defining this pair of fields as variable length. This then would be interpreted, by the computer, in such a way that the computer would look to the 5-bit fixed field and find out how many characters were contained in that occurrence of the related data field over and above 27. The computer would then process the related data field based on detailed knowledge of its instantaneous length.

A variant of the variable field concept is the specified field length. If, by detailed statistical analysis of the file, the programmer were able to determine that the length of a field varied significantly, but not randomly, then the programmer would consider an alternative field definition. Such a field definition would be termed "specified." A field would be specified as to length if successive entries had fields whose length were precisely the same as the immediately preceding field of the same name. A programmer would analyze the file looking for runs of the same field length. If this phenomenon existed, he would introduce a new entry type whose sole purpose would be to change the length specification for a specific data field.

The first entry in the file would be an entry that set the field length to its initial value. Following this would come entries of an alternate type that contained data. When an entry occurred whose length did not conform to the existing specification, a field-length change would be introduced that would amend the specified field length or replace it by a new specified length. An entry of the new length could then occur. Additional entries would be allowed until it was necessary to change the length specification again.

Some Examples of Data Fields

To avoid getting too far afield let us test the definitions of these three important terms—fixed, specified, and variable—with examples from our personal experience. Once a library is established, the date of such establishment is fixed in all of its characteristics. The number of employees on duty at any hour of the day is specified. The amount of the budget remaining at the end of the month must be variable.

The same three powerful terms can be used to help describe other characteristics of a data field. The character set used for a Library of Congress catalog card number is fixed. The first two digits of the Library of Congress card number are specified for a 12-month period. The length of a title on a new catalog card is variable. (While computer people might like to take credit for the above concepts, unfortunately they cannot. The figures-vs.-letters shift on a normal teletype is a technique for specifying the character set to be used. In this way the basket of the teletype is latched into one 5-bit (31 characters) set or the alternate character set. This, in teletype messages, is a method of specifying the character set to be used until the specification is altered.) The specified mode is an attribute of serial files and serial storage media.

In a similar manner, most punched card work has been done in fixed length fields with a fixed character set. While this is not a limitation of the media (until the physical limitation of 80 columns has been attained), it was a limitation of the devices that were processing such cards.

Again to pick an example from the communications industry, the length of a teletype message is limited, first by the roll of paper tape on the machine and second by the end-of-message indication. Teletype messages can be a variable number of lines in length. The lines themselves can vary in length. To describe such a teletype message to a computer, one would either define the end-of-line and end-of-message character encodings so that the computer could scan line and message length itself, or a fixed length field would be associated with each line and would give the number of characters in that line. The whole message would be preceded by a total line count. These counts could either be imbedded in the message or gathered together at the head or the tail.

One may reasonably ask, why give all this overly detailed attention to character sets, field length, and encoding? There are two reasons. First, this material is well known to a few senior people in the data processing field but has never before appeared in print. Second, the file storage units will be a significant portion of the cost of an automated library system.

The total volume of the information to be stored can be altered as much as 100 percent by the proper application of the above techniques. Therefore, the total cost of computer equipment for an automated library can vary, depending on how well the above techniques are applied. In the case of the National Union Catalog, this variance is approximately $5 million.

To determine which of the above techniques is appropriate, a file designer would examine *each* field and compute the number of bits required to contain all of the occurrences of that field if it were fixed, if it varied and had a count field appended, and if it were specified and the specification changed with each transaction. Given intimate knowledge of the file, the one "best" description can be determined by minimizing the bits to be stored.

One point of encouragement here: when one dealt with inflexible media such as punched cards or typeset cards, the format, description, and specification of each field was inalterably set at the time of initial design. This caused one to be ultraconservative in the initial description of the file. For the same reason this conservatism is evident in the design and ordering of our present card catalogs. The die, once cast, is permanent. With magnetic media such is not the case.

It would be unfair to paint too rosy a picture but several facts may be observed. The file will never be static and hence will be updated and maintained regularly. By careful changes to the maintenance program the specifications for the file can be altered and a restructured file obtained as a byproduct of its regular maintenance. Thus, if a field were called fixed, and eventually grew out of the established bounds, it could be redefined (by a careful alteration of the computer programs) before one of the regular maintenance passes.

Selection and Searches

As intimated above, a file is kept in some order determined by specifying one or more data fields to be used as a sort key. When more than one data field is specified to be used as a sort key, they are combined temporarily into a superfield. As sequential entries are obtained, the sort keys are compared and pairs of entries reordered until the desired sequence is obtained for all entries.

If a file were to be placed in ascending alphabetical order by author, these fields—last name, first name, middle name, and year of birth—would be defined as the sort key. Their combination would be specified in the order given. The year of birth would be defined as a fixed-length numeric field, whereas all others would be defined as variable-length alphabetic fields. The variance in their length would be described by a leading character count or by a terminal punctuation mark, whichever suited the particular computer best. In this manner the desired sequencing would be obtained; the principal modification required to bring this simple example into contact with reality would be to expand the file whenever multiple authors appeared, using a permutation scheme similar to that used in KWIC.[6]

Earlier we defined the selection operation. We may now attempt a more precise definition of this important function. The simplest use of an index file is to reference it, by man or machine, to obtain the location of an item of information (document, report, monograph, or serial) in the collection. *When* the document can be unambiguously described by the requester, *and* if the fields given by the requester are the same fields on which the file is ordered, *and* if the request is unique, *then* a direct selection can be made and the whereabouts of the document determined.

For example, if a reference librarian knew the author and the full title of the document, and if the index file were ordered alphabetically on author and then on title and, furthermore, contained either the shelf location of the document, or a record of where it might be borrowed, or the person to whom it was now charged, a direct selection could be made from the file. The results of this selection would either be a call slip for the clerk in the stack, a form letter to a cooperating library requesting a loan, or the name and phone number of the person currently charged with the document.

The other interesting function performed on files is a search. If a requester ambiguously describes a document, or if the file is not ordered on the data fields supplied with the request, a search is required. There must be a criterion for initiating the search, a criterion for terminating the search, and a criterion for success. These three criteria may be as simple or as complex as allowed by the file designer and the computer programmer. As might be expected, the more the request deviates from the key on which the file is ordered, the more arduous the search will be. The worst case is, of

[6] Key Word In Context. One of several terms used to describe permutation indexing techniques.

course, the complete passage of the entire file on some obscure combination of data fields.

A trivial example of such an obscure search would be a request for an illustrated book of odd size printed in England in 1955, dealing with the medieval period. For such a request a full search of the index file would be required, unless a current file in order by subject were available. If an index file were retained in subject order, a search still would be required of the section, *medieval history*. The search would be initiated at the beginning (or end) of the section on medieval history and would be terminated by the first occurrence of a book referring to another period of history. During the search, each entry would be inspected to see whether the subject document were illustrated, of odd or standard size, and printed around 1955. Obviously, searches can be performed on all data fields within an entry and on any combination of quantities contained within any combination of data fields. If the author's last name and one or more keywords in the title were known, only a restricted search would be required.

Hardware Specifications

In recent history the only practical storage medium for voluminous files has been magnetic tape. Typically, this tape comes in 2,400-foot spools, contains 6 binary bits per frame, and approximately 1,000 frames per inch. Information is stored by passing the tape under a set of electromagnetic heads and creating spots of local magnetism on the tape at the aforementioned density while the tape is moving 150 inches per second.

Information is deposited on tape in variable-length blocks called "physical records." The physical records are separated by a ¾-inch gap of erased tape. Tape units are constructed so that they can stop and start again within this small gap. The computer reads or writes one physical record at a time. Records may be passed over in either direction without altering them.

By applying standard usage factors, one finds that such a tape would hold approximately 5,560 physical records of 24,000 bits each. Thus, about 134 million bits of information could be stored on a magnetic tape. This information could be passed under the reading head in 176 seconds—a shade under 3 minutes. It has been estimated that a typical catalog card contains 1,000 bits of information. (See item 7.) Thus a linear search of 134,000 catalog cards could be performed in slightly over 3 minutes.

Modern computers have their input/output organized in a manner that is described as buffered. This means that the computer can perform other operations completely independent of input/output transactions.

Modern computers can have several buffered channels. An information rate such as that described above is possible over each of the channels with complete simultaneity, provided the computer is fast enough to digest the information supplied to it before the next information is presented. A good-sized computer will allow 8 input or output channels to be operating concurrently. Thus, the equivalent of slightly over a million catalog cards could be searched in 3 minutes. This type of search would not be required very often, but it is important to know that present commercially available computers could handle these search volumes in such a reasonable period of time. This is a far cry from the usual pathetic computer, which is pictured performing a linear search of the entire file system in serial fashion.

Before we leave the subject of linear tape searches, one more obvious observation must be made. By the simple expedient of a table containing the first and last entry on a tape file, the length of search can quite often be diminished. In the aforementioned case, where author and title keywords were known, the tape spools that preceded the one containing the author's name would never be searched. Spools containing names higher in alphabetical sequence than the name of interest would similarly be bypassed. For the trivial case mentioned previously, or for a search on the number of illustrations, type font, and birthplace of author, a serial tape search would be required for all items in the catalog.

The amount of information contained on one spool of tape and hence the length of time required to pass that spool of tape are related, in an almost linear fashion, to the length of the average individual entry. Previously we discussed in detail how fields could be defined as fixed, specified, or variable, in order to minimize the total file volume required for the storage of an average entry. Now we should like to discuss further the concept of "split vs. combined files" as an additional technique for reducing the volume of file storage while simultaneously reducing the search time.

This concept offers an improvement in addition to that previously discussed, i.e. minimizing the length of the average record. If a file is to be stored in one order only, the search time may be reduced by splitting the file into potential search keys and respondents to such a search.

Clearly, searches will always be allowed on author, title, classification, and keyword. It is a moot question whether to allow searches on the author's birth or death dates, the illustration statement, or the publisher. Most librarians would agree that color or style of binding, size, number of pages, and copy number are, instead, respondents to a search along with location and security classification, if any. Thus it is possible to split the file into potential search keys and improbable search keys.

The potential search keys would be contained in a search entry. Such a file would consist of entries having a minimum average record length and hence would search as rapidly as possible. Along with the keys in each search entry would be the location (address) of the full response to the search. The shortened (compact) file could be searched and the addresses of the respondents retained within the computer in a list. After such a search, the response tape(s) could be placed on the machine and the full response could be selected and printed out. With magnetic tapes, such a two-step process requires batching for economic reasons. Batching increases the turnaround time for a request while gaining efficiency in the operation.

If the file is to be retained in more than one order (trading storage space for search time), the split file is even more attractive. Rather than retain the entire file in every order, only search entries are kept in every order, while the response file, containing complete bibliographic data, is in only one order. All potential responses are batched against the response file in a subsequent operation.

We have been talking of familiar concepts in unfamiliar terms. The catalog card is itself an entry in an index file. The lines on it vary in length and number within a physical maximum. The index file is kept in alphabetical order on one or more fields. Sometimes a file is duplicated and retained separately by both subject and author. For example, if there is only one subject entry and one author for each item represented in the file, the total volume stored is double that of either file singly, but the information content is the same. The pair of files are doubly maintained in order that search time can be reduced at the expense of file volume.

Sometimes these two files are merged into a double-length file in combined author/subject order. Humans make selections from these files or embark upon limited linear searches. In every case the drawer labels are used to block out a search just as the labels on the front of a magnetic tape would be used in a decision process either to search or block out a tape.

In the last 2 years practical magnetic-disk stores for computers have trickled into general usage. Magnetic disks and drums have been under consideration for a long time, but only recently has their capacity and reliability placed them in the practical category for large volume storage. However, the above definitions and discussions were not wasted since they have allowed us to understand the relatively simple serial media and have prepared us for a discourse on a cyclic storage media.

We must apologize for a misnomer in daily use in the data processing field. The term "serial" is applied to tapes of all kinds and implies that search time is a function of the length of an average entry and the number of entries searched. Conversely, "random" is usually applied to magnetic drum and disk. This implies that the delay in accessing an item of data is not position dependent. Such is, unfortunately, not the case.

In the case of the common magnetic drum, the electromagnetic heads for reading information are fixed to the frame of the machine and the magnetic media rotate on the periphery of a steel cylinder, which is driven about its axis of rotation. The access to any spot on the surface of the cylinder (drum) is a function of the relative position of the head (fixed to the machine frame) and the spot in question (rotating with the surface). Such storage media should be described as cyclic.

Further to confound the initiate, the magnetic-disk storage device was developed.[7] This is simply

[7] A similar pattern of development occurred with the phonograph. Originally cylinders (drums) were used, then disks.

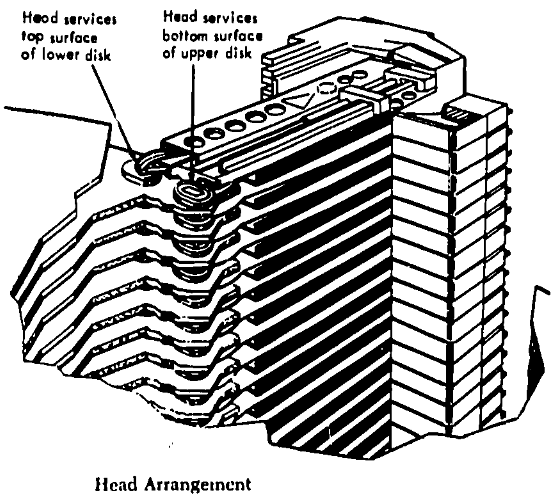


FIGURE 3. *Arrangement of read/write heads, disk storage.*

a technique to obtain more area for the deposition of magnetic material. A series of circular disks is bolted to a single shaft and that shaft is rotated at constant speed. Magnetic material is placed on both surfaces of the disk. The read/write heads (see fig. 3) are servo-positioned to a specified radius under computer control. There are usually one or more heads per disk surface. Once the heads are positioned, the access to concentric circles of information (called tracks) is cyclic as in the case of the drum.

In addition, tracks on other disk surfaces at the same radius may be read (as in the case of the drum) by switching electronically to alternate heads, which have already been positioned (see fig. 4). Access to other tracks at other radii is obtained by payment of an additional time penalty (measured in milliseconds) used for the positioning of the servo arm, which carries the heads themselves. Thus a magnetic disk is far from random but is both current-position dependent *and* cyclic.

We ask ourselves: Why is all this mechanism necessary and what does it gain us? Storage capacity and ease of operation are the answers. A magnetic disk is relatively inexpensive and is sealed in a protective environment, which immediately avoids the usual dust problem with magnetic tape and, therefore, increases the reliability.

In the previous example on magnetic tapes, we pointed out that approximately a million catalog cards could be scanned in 3 minutes. The question that quite naturally occurred to most of you was: What if I have more than a million cards? The answer is: Change tapes! This is something not normally mentioned. A tape change, by a competent operator, takes approximately 2 minutes per unit. If 2 million cards were to be scanned in the previous example, 8 operators would be required to perform this in the minimum time and 8 minutes (3+2+3) would be required for the search. If less than 8 operators were available or if the operators were not alert enough to perform the tape

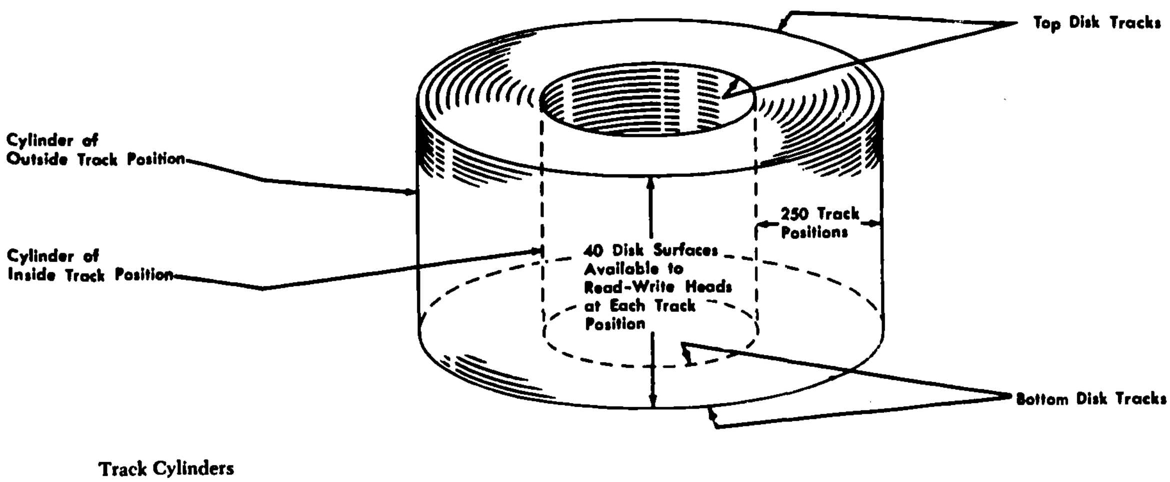


FIGURE 4. *Magnetic disk storage.*

735-898 O—64——4

change in the minimum time, then 8 minutes would not suffice. If the National Union Catalog is conconsidered, then 14 x 5 or 70 minutes would be required for a single search if perfect operation were presumed. Clearly this is so impractical that it would never be contemplated. On the other hand, a single magnetic-disk storage cabinet of the kind presently available contains 336 million bits. Those expected in the reasonable future should contain 650 million bits. If the 14 million cards of the National Union Catalog are to be stored at 1,000 bits each, then 14 billion bits of storage would be required. Twenty-two of the 650-million-bit cabinets would be needed to store this file.

Though formerly the utility of split files was not immediately obvious, we can now see their benefit. If the file were split into request keys and respondents, short subfiles would result. Such subfiles could be searched at a rate of 900,000 bits per second per channel. A specific place on each disk could be set aside for tables that define the starting key and the stopping key for each cylinder, much as the alphabetical card on the front of a catalog drawer now does. Access could be had to any block of 2,600 cards in approximately 100 milliseconds and these 2,600 cards could be searched in about 1.55 seconds. Some of the techniques developed for block searching and block sorting could be applied to such a mechanized index file, and hence great speed of search could result.

Search speed, *per se*, is not necessarily valuable. But search speed in the presence of sufficient demand yields a low cost per request. The storage device described above would:

1. Rotate every 34 milliseconds
2. Contain a read/write head for each disk surface
3. Contain some 40 surfaces
4. Store 650 million bits per cabinet
5. Read or write at 900,000 bits per second per channel
6. Have 10,000 tracks per cabinet
7. Position the heads from track to track in approximately 100 milliseconds
8. Cost about $250,000
9. Store a bit for 6/10 of a mill ($.0006)

One should not make the mistake of comparing this storage price directly with magnetic tape. Disk storage devices are usually permanently dedicated to the file they contain. On the contrary, magnetic tape is removable and the stored media only cost about $50 per spool. The mechanical tape-transport unit and its associated electronics can then be used for other purposes. Again, as in most computer undertakings, there is a tradeoff of cost vs. time.

Some Observations on Information Retrieval Schemes

We are now in the position to make some interesting observations and judgments on information retrieval schemes. The coordinate file is what we now know as a combined file. We realize that this is excellent for selections, but, since the average entry length is long, it is not very good for linear searches and is particularly wasteful of storage space if the complete file is to be retained in several orders. On the other hand, the inverted file makes use of the columns described in Vickery's "information array," (see item 8) and, furthermore, is a split file. A coordinate (master) file also exists somewhere but the keywords have been split into subfiles, collected, and compressed, so that their redundant information is removed. (See the appendix to this paper.)

An inverted file is an instance of a split file that makes use of two entry types and field specifications. Instead of specifying the length of a field, the alternate entry type of an inverted file specifies the keyword (descriptor)[8] that applies until altered by an entry that establishes a new outstanding descriptor. While a descriptor is established and outstanding, the file consists solely of item numbers which are associated with that descriptor. After lists of items associated with all the descriptors are obtained, they are matched with one another until the proper combination of AND, OR, ALL, and NOT is obtained. The inverted file with its nonredundant entries is a phenomenon of serial storage media, particularly magnetic tape (or magnetic disk when used in cylinder mode).

Both types of file organizations are outgrowths of previously kept, manual index files and were not developed specifically for machine use. The file organization given in the appendix to this paper

[8] Descriptors may be of any type, e.g., subject, terms, dates, names, etc.

uses the definitions and distinctions made thus far, states its assumptions, and provides some parametric measures whereby its applicability can be measured. This organization is, to the knowledge of the authors, unique and one of the first designed specifically for the machine searching of files held on magnetic-disk storages.

File Conversion at the Source

Now that files have been defined and described, let us continue with our analysis of the conversion problem. In the first half of this paper, we have followed the lead of the many eminent authors who preceded us by assuming files into existence and then hypothesizing how they might be used, maintained, and referenced. If we probe deeper, we must eventually come face to face with the question: How do these files initially come into being? Let us hasten to agree with other wee voices who have cried in the wilderness and state that the economic, optimum, practical, recommended, and most logical place to automate these basic files is at the source. (See items 5 and 6.) Clearly we are remiss if we do not obtain a machine-readable record as a byproduct of the original cataloging operation.

There are several ways in which this might be done. Any or all of them could be applied. We cannot suggest strongly enough that some one of these excellent techniques be applied immediately, posthaste. A punched paper tape could be readily obtained as the byproduct of any of the early key-driven operations whereby a catalog card is produced. Several commercial newspapers are obtaining a punched paper tape as a byproduct of the story copy produced by the staff reporter. During the editing process the hard-copy sheet is marked up. The editor's marks are then encoded and a subsequent computer run merges the editor's marks, corrections, and deletions with the previously punched paper tape. The resulting updated copy is then sent directly to the printer. With the availability of machines such as those discussed in the paper on output printing, characters for typesetting control can be interspersed automatically by computer and a catalog card can result. If such machines are available, there is nothing to hinder this process now—others are doing it; the equipment exists; libraries merely are not benefiting. Adopting such a process could result in savings of both cost and time for libraries.

On the other hand, the typewriters used by the catalogers could be obtained with the Farrington Selfchek font and the preparation process could be continued just as it is now. Then, once a catalog card had been completely prepared, the card itself could be optically scanned and a "magic" typesetter used to produce the cards. This process would also offer economy, speed, and efficiency. The machines exist now; nothing is stopping us except ourselves and, perhaps, the funds.

If either of the above techniques were to be adopted, the magnitude of the unreadable National Union Catalog would remain constant. As it is, the catalog now stands at about 14 million cards and is increasing by approximately 1 percent every year. Although 1 percent does not seem to be excessive, a 5-year accumulation of new acquisitions is 700,000 catalog cards! So much for what can do to reduce our burden in the future. But what of the burden we have now?

File Errors, Editing, and Conversion

Files may be classified as recirculating or reference. A recirculating file experiences 100 percent activity in a reasonably short time. An example of such would be an insurance file, which recirculates either on the month, quarter, half, or year.

The National Union Catalog is a reference file. The activity in a reference file is so low that only a small portion of it is ever referenced. Since the activity on a reference file is low, the file usually grows monotonically. It is never purged. On the other hand, a recirculating file is purged frequently, at least once every cycle.

Another interesting distinction can be made between files on the basis of their accuracy. A clean file is a collection of entries, each of which was precisely correct at the time of its inclusion in the file. On the other hand, a dirty file is a file that contains a significant portion of errors. A recirculating file is purged and cleansed as it cycles—a utility-company billing file is of this nature. After the file "settles down," the proportion of errors imbedded in the file is a function of the new activity applied to the file. The error rate is normalized with respect to the business cycle.

In a large reference file, errors are something merely to be contended with. Systematic errors usually are worked out, but random errors are trivial inconsistencies that usually are never corrected. In addition, a large reference file undergoes evolutionary changes apace with the need and the requirements for such a file. New specifications are reflected almost immediately in the new activity. The overseers of such a file have every intention of bringing all of the previous entries up to the new standard, but often these intentions are never translated into action.

Before librarians picture the National Union Catalog and feel sorry for themselves, they should note that others have similar problems. Law enforcement files are also noncirculating reference files. They grow monotonically from information that was incomplete and possibly incorrect initially. Furthermore, there is no good way to purge such files, since men do not, as their last act, process a purging transaction against their own entry.

Federal census files are also of this ilk. While the census does get a new lease on life every 10 years, these files are only statistically accurate; their individual entries are subject to error since they are not verified with the individual involved. An important factor determines, to a great extent, the cleanliness of a file. If no feedback is provided to the appropriate individual, or if the appropriate individual does not care, a file is inclined to be error prone. In the case of the National Union Catalog, if the preliminary catalog card does not carry the cataloger's initials before it is finally typeset, the feedback path has not been completed and the file is error prone. Multiple checking, verifying (the same operation repeated by a different person), redundancy, and proofreading are all excellent clerical techniques to reduce errors, but they do not eliminate them. It is estimated that the National Union Catalog contains a significant amount of errors, as high as 5 percent.

As a springboard for discussion let us assume that computer editing could locate three errors out of every five and flag them for correction. The problem of coping with the remaining 2 percent (overall) errors is not purely academic; 2 percent of this catalog is approximately 250,000 entries. We must ask ourselves if this is too much error to tolerate and, if the answer is yes, how much we are willing to spend to purge it. As indicated below, this greatly affects our conversion technique.

The initial conversion of a large file is one of the most underestimated automation tasks. If one is automating a file that does not recirculate and cleanse itself, there is no way of knowing the initial state of the file without performing a pilot study with a statistically meaningful sample. Again and again, our military counterparts underestimate the magnitude of the job. This is invariably the cause for a major change in scope, considerably more work on the part of the contractor, and perhaps additional contract negotiations over money. As mentioned before, without feedback, there is no way for a file to cleanse itself.

To be sure, all of the people who handle our present manual files are well-motivated, loyal employees. But they do make mistakes and, in some cases, our most senior people are not assigned file maintenance tasks.

The errors in our present files can be classified into two categories: filing errors, where manual interfiling is incorrectly done; and source errors, where one or more fields in the data entry itself are incorrect. Of course, the two types of errors are not completely independent but for the purposes of this discussion can be considered so.

While filing errors are a great handicap in a manual system, such errors can be automatically recognized and corrected once the file has been mechanized. In computer parlance, a program would "sequence check" the file. If an out-of-sequence condition were to occur, the suspected entry and the entries immediately preceding and succeeding it would be printed for human review. Eventually all out-of-sequence errors would be corrected and only source errors would remain.

Source errors may be eliminated in either of two ways. Through the mechanism of feedback, the catalog card itself can be checked by a knowledgeable individual, different from the one who made up the card initially, and errors thus located and eliminated. The second person may be either the author of the work cataloged, another cataloger, or both.

The other technique for catching errors in the source documents themselves is to have a highly motivated person study the file long enough to become intimately acquainted with it, and then, in the course of the study, point out any discrepancies as apparent errors. To perform this second task, rules defining the formats must be rigorously described. These rules would define each entry type and all of the allowable field combinations within each entry type. Given a complete set of definitions an individual can carefully peruse the file and flag any unusual occurrences for further study.

Manual Conversion Techniques

The most popular means for manually converting a clean file is the keypunch. If a file is already clean, this purity should be retained during the conversion operation. This is usually accomplished in the following way. An analyst devises a card form and describes in detail the rules for its use. Clerks (usually female) are trained in the use of an electromechanical device called a keypunch. This device, manufactured by IBM, operates in the following way. It has a supply of unpunched cards having space for 80 columns. In response to key strokes the cards are fed from the hopper across the bed of the machine and into the stacker. As they pass a punching station, the operator visually reads from the hard copy and records the characters she has just read with key strokes; these key strokes are transmitted electromechanically to a set of punch dies which pierce holes in the card.

The normal keypunch has 47 separate key strokes, which result in 47 discrete unambiguous hole combinations in the card. A special device can be ordered, at a slight additional cost, that raises the allowable number of combinations to 64. The number of characters in the basic character set is of little importance to us, since, in the original design of the input-card form, a mode-change character that is used to select a different character set from the normal can be specified. By a sequence of mode-change characters, as many character sets, type fonts, or alphabets as are required may be had. This is merely an adaptation of the mode-change character commonly found in teletype communications; however, since the advent of the computer, more generality can be attached to this simple idea.

The character set initially specified is the natural character set of the keypunch. That specification remains intact until a new specification (immediately following a mode-change character) is set. In the following rather simple way, a catalog card could be easily punched: start with a mode-change character that set the mode to boldface alphabetic; following this would come the author's name; following the trailing punctuation would be a mode-change character that specified the mode as numeric; the author's year of birth would follow; whenever this line was complete, an end-of-the-line symbol would establish the format for the title line; and so on down the card, with the changes of font, capitalization, intervening punctuation, through the end of the recorded information. If the card were multilingual (e.g. Russian-English), a mode-change character would merely indicate this fact and the punching could continue in Cyrillic or any other alphabet.

The keypunching operation proceeds until a clerk has completed a batch of source cards. When this has occurred, the batch of source cards and the resulting deck of punched cards are transmitted to a second operator (different from the first) who utilizes a similar machine called a verifier. The verifier has a physical appearance similar to the keypunch.

During the verifying operation the operator depresses keys as before, but instead of the keys controlling piercing dies, they control a pattern of sensing pins. If the pattern of holes already punched in the card matches the pattern of sensing pins, that one character is considered verified. The card is then advanced to the next card column. If the pattern of pins and the pattern of holes do not agree, an alert is set that must be cleared by manual action. In this manner, the cleanliness of the original file can be retained.

It is estimated that three 80-column punched cards would be required to hold the information contained on a single catalog card. It is usually assumed that a punched and verified card costs 10 cents. The three punched cards that contain the same information as 1 card from the National Union Catalog would cost 30 cents. By this means, the first 10 million NUC cards would cost $3 million to transcribe from their present form into a machine-readable form. This sum includes the cost of the card stock, the rental of the machines, the salaries of the clerks and their supervisors, and the overhead for this task force.

It is interesting to note that a byproduct of the punching and verifying operation is the editorial review described above. Keypunch operators can become extremely familiar with the basic structure of the file and are quite adept at flagging apparent errors for further study.

A common alternative manual technique substitutes rolls of paper tape for the separate punched cards. If the verifying operation is not required, then a device similar to an electric typewriter is used.[9] This device produces both a hard copy and a roll of perforated paper tape. The same encoding schemes described above still apply; holes are perforated in the paper tape in response to the operator's key strokes. The visual listing is used for dynamic visual verifying by the same operator who performs the key strokes. The quality of the check is good, though not as good as if a second operator had performed the verification.

Punched paper tape verifiers have been used, but they have difficulty, as one might surmise, when an error is found during the verifying operation. Unless automatic repunching (with a second punch attached) is used, the only choice the operator has is to null out the field in error and process a cleanup transaction later. This is due to the serial media—the tape following the error is already perforated and the error cannot be easily corrected where it stands. If the file is not clean enough to warrant verification by a second operator, the paper-tape system is excellent since it provides a hard-copy byproduct for visual verification as the tape perforation is taking place.

Both of these schemes have the advantage that a well-trained clerk is required to read the source documents, word by word, and stroke the keys, character by character. It is this detailed operation that provides a person with the intimate contact necessary to recognize and flag apparent errors in the source data.

Automatic Conversion Techniques

Another conversion technique is optical scanning, a rather new technique that is just beginning to have extensive use. The majority of the successes with optical scanning are credited to Farrington Electronics, Inc., of Alexandria, Virginia. They have constructed 100 machines that will scan 1 or 2 lines from a card or many lines from a typewritten page. At the present time these machines are constructed with a single character set of 64 characters. To increase the probability of a successful read, Farrington recommends that its customers use a special type font known as Selfchek. This type font can be had for almost all electric typewriters and, in addition to being rather easily read by the optical readers, is pleasing to the eye. At the present time the document handlers for Farrington's standard line of equipment accept either punched-card-size documents or page-size documents. There is nothing in the scheme that prohibits the handling of catalog cards, although this would be a special order.

For the purposes of converting the National Union Catalog to magnetic tape, these devices have two limitations as they are presently constituted. They have only one character set, and they are devised as completely off-line devices. That is, there is no provision for editing (as the girl at the keypunch performs it) in the basic devices. It does appear as though these devices could be expanded to handle more than 64 characters, but this would be a special development. Since Farrington does not have, at present, a multifont optical scanner in production, and since really heavy use of these devices is still in the future, some assumptions were necessary in order that a ball-park figure for the mechanical conversion could be devised.

After these assumptions were made, one concludes that present state-of-the-art optical equipment could process a catalog card for about 3 cents. For the purpose of deriving a rough estimate, say that the National Union Catalog has about 10 million cards to be converted and that only 90 percent of these cards could be processed optically. Conversion for these 9 million cards would cost 3 cents a card, or $270,000. A million cards would remain to be keypunched. At our estimate of 30 cents a card for keypunching, this would amount to $300,000. Thus a 10-million-card catalog could be converted for $570,000.

To avoid the real possibility of being quoted out of context, the assumptions in the above costs must be stated. It was assumed that 90 percent of the National Union Catalog could be read by an extended optical page scanner. It was assumed that some other editing technique would be judged sufficient to perform the detailed scrutiny

[9] Three such commercial devices are: the common teletype with punch attachment, the Flexowriter, and the Dura Mach 10.

discussed above. It was also assumed that the optical devices would run 20 hours per day with little additional maintenance and that the remaining 10 percent of the card catalog would be keypunched and verified.

It should be remembered that no techniques can be recommended at this time since the task to be accomplished has not yet been defined. Further, it should be remembered that we are not recommending these techniques; they are merely described in terms of presently available hardware as a guide to your thinking.

Semiautomatic Conversion by Stenotypy

An additional technique, sometimes mentioned, involves the use of stenographic recorders, similar to those used by court reporters, and an optical scanner to read the tapes into the computer. The stenotypist would record abbreviations on the stenotape of what she visually read. A computer would read this tape optically and, during the input reading process, enter into a large dictionary of abbreviations in order to retrieve the full natural-language spelling for the abbreviated word. The abbreviation would be replaced by the full word before the information was stored in the file.

If a data file is extremely redundant (such as a natural-language text) then the stenotyping technique cuts the number of key strokes required by as much as two-thirds. This technique can be three times faster than normal key stroking with resultant savings. The National Union Catalog contains a minimum amount of redundancy. Whenever numbers are involved, there is little or no redundancy; the same holds true for names and most titles. Thus, while the stenotype technique would be appropriate for the conversion of files of text, it is not especially attractive for catalog card conversion.

We have three techniques for converting a file to machine-readable form. They are: 1) keypunch and verify with manual scrutiny, at a cost of 30 cents per catalog card; 2) retype with punched paper tape and visual hard-copy editing by the same operator, at a cost of approximately 15 cents per catalog card; 3) optical scanning, with limited multifont capability with no manual operation and no checking, at a cost of approximately 6 cents per card. Clearly the optical scanning shows great promise and should be investigated in detail.

Computer Editing of the Converted File

After the file has been converted to machine-readable form, the work in producing a usable index file has just begun. While much of the work lies before us, most of the really hard labor is past. The file now resides in a machine-manipulatable form, probably magnetic tape. Through a series of computer passes, this magnetic tape will be repeatedly manipulated and a final index file will result. The first operation performed is a field-by-field edit of each file entry to determine if the data received by the computer adheres to the limitations set down in the data description for that field. This first, and most primitive, edit pass might perform, for example, a check to see that only alphabetic information plus limited punctuation appeared in the field called author's name.[10] In a similar manner the location of the document field would be edited for legal locations. This type of edit would be classified as a "character-set edit by field definition."

Usually more than one computer program is required to edit a file. These programs are progressive in nature. The file is repeatedly processed by the sequence of progressive computer programs until it has attained a sufficient degree of polish. After the errors found in the first edit pass are removed from the file, another edit pass is made over the file and, through the use of context, illegal combinations of entries are located and flagged for elimination. As the computer analysts gain more familiarity with the file, they devise more sophisticated editing techniques so that the more subtle errors can be located and found. An example of a more sophisticated editing question is the relationship of the publication date of the document and the year of the author's birth.

As a byproduct of the above edit passes, blank fields will be located and flagged for manual action. This flagging would consist of printing out the entry identification, its location in the file, and a description of the omission. After subsequent manual action, a transaction would be posted against the file either to complete the entry or to fill the void with a recognizable null character. Each entry will be polished until it meets at least the minimum applicable edit criteria.

[10] There are exceptions to this which every librarian will recognize, but which need not concern us here.

As was mentioned previously, no specifications exist for how error free a catalog file should be, nor do we have any indications as to how much money we are willing to pay for the information in a mechanized catalog file. One example will suffice to show the problem involved. It is traditional that the entry for a personal author usually has a birth date and, if he is deceased, a year of demise. A trivial computer program could read the file and list all of the authors for whom date of birth was not known. Research librarians could determine the date of birth of all of these individuals or flag the computer that this was an author entry which did not require year of birth, so that the appropriate notation, either way, could be made in the file. As a byproduct of this posting transaction, the research librarian could contribute any dates about deceased individuals he found conveniently available.

In a subsequent computer run, the computer could print out the names of all authors with no date of death and who would now be more than 100 years old. The task before the research librarian is clear. Again, more dates are processed to the file. After repeated iterations of this process, the file would be relatively clean regarding the period in which the author lived. All that remains then is to establish a technique whereby an author notifies the Library of Congress as he breathes his last breath: then the file could always be current.

We chose this data field for an example because of the frivolity involved in such a venture. In these days of limited resources, our money can be better spent. The case for indexing in depth is not quite as clear cut. The computer could be asked to list all of the entries that do not possess at least five keywords. As the catalogers went through this huge printout, some entries would be noted for additional indexing and rework. Others would be flagged complete as they stood. Over and over the process could continue until the available resources for this purpose were expended.

These are but two examples to indicate the way a computer might be used to edit for discrepancies between the data field and its description, or for omissions from the file. Other contextual editing programs could attack the problems of identical names, near-names (two spellings so close that this may be the same author or title), duplicate entries or volumes missing from a series. The list is unending. The cost is magnificent.

File Changes, Additions, and Deletions

Before any conversion operation is contemplated, the system designers must face the problem of the transition period. How are changes, additions, and deletions to be handled? Using two optical scanners (described above), the National Union Catalog could be converted in approximately a calendar year of three-shift operation. During that time, however, 120,000 new catalog cards would have been received and posted to the file. Changes are not really as difficult a problem for a mechanized file as one might anticipate. One merely makes a copy of every transaction posted to the manual file, starting on the day the conversion operation is initiated. The extra copy is stored in chronological order in a separate collection of changes, additions, and deletions.

After the main file is converted, the subfile of changes, additions, and deletions is also converted. The subfile is sorted to the same order as the master file and a simple update operation, similar to that used in maintaining payroll files, takes place. The new transaction takes precedence over the old, and a completely new, updated master file is obtained. The same update operation will be required after the mechanized system is operational. The first day's transactions pick up all the changes to date. From this time on the file cycles normally.

An incomplete entry is replaced by a complete entry. An erroneous entry is replaced by a completely new, correct entry. A previously nonexistent entry is added to the tape file. If, for some reason, a card not to be replaced by a new, corrected card is removed from the catalog, then the card removed is sent to the subfile. It is processed as a straight deletion with a separate transaction code.

The problem of changes, additions, and deletions is not a significant one with an automated file. The administrative procedures are rigorously enforced so that all of the modifications are captured in the change file. The appropriate changes are reflected in the file in one overnight operation before the file becomes operational.

The Master File

Earlier we implied that all of the information on the National Union Catalog card might be converted to magnetic tape. To be completely clear, the following was meant: whenever a boldface was encountered, this would be recorded magnetically. The number of lines and the length of each line would be recorded. The number of separate type fonts and those words utilizing them also would be recorded. The file would contain information as to what words had been in boldface, Cyrillic, italics, etc. The entire file would be converted, and the information from the entire file would be available. Then if it ever became necessary, phototypesetting devices could be computer driven and the original catalog cards reproduced. Obviously, if console displays did not require all of the format information available, a lesser entry could be output.

All of the information from the initial input would be retained in one master entry until the aforementioned edit operations were completed. When the time came for the automatic library to begin to function, the file would be split. All of the information concerning the format and type fonts on the original catalog card would be placed in a subfile. Of the remaining information, some would be placed in an author file and sequenced on author's name. In the case of multiple authors the same title would appear more than once. In a similar manner, a title file would be split off and kept in sort by title. Lastly, a subject file might be kept in some special search order similar to that given in the appendix to this paper.

Special index files would be maintained by author and title to allow direct selections if the requester knew the identity of the document. In addition, one or more index files would be kept for searching. In each case, an entry would probably not be complete without a reference to the master file. It should be noted that the master file is never searched but is only used as the object of a selection operation.

Compression and Packing of Files

One axiom of the computer field states that you can always trade time for space. Nowhere is this more true than in the area of file design. Each modern digital computer has some natural unit of information. In some machines this is called a character and the machine naturally handles 6 bits at a time. In other machines this is called a word, and the machine naturally handles either 36 or 48 bits at a time. Machines operate their fastest when a file is designed so that the individual information fields are contained in one or more of these natural lengths. If one of the fields in an entry were "number of authors" and provision were made for holding 16 or less authors, this could be handled in the most expeditious fashion by placing it in a full computer word. If we assumed for the purposes of discussion that a computer word were 36 bits in length, then there would be a waste of 32 bit positions if a full word were awarded this purpose. In other cases the waste may not be so spectacular. The extent of compression possible usually hovers around 50 percent: an unpacked file is almost twice the length of a packed one! [11]

Most file designers consider putting several short fields together in one computer word to gain efficient utilization of storage. It should be noted that although the storage is efficiently utilized additional computer time is required for the unpacking of these fields before use. Many computers have special instructions in their repertory to facilitate the packing and unpacking operations.

For very large files, such as the National Union Catalog, second-order packing is frequently done. As a prerequisite to second-order packing, the file designer needs to be extremely conversant with each of the data fields in an entry. He makes use, wherever possible, of some phenomenon peculiar to a data field. This is easiest to appreciate when numeric fields are considered. Consider, for example, the year of an author's birth.[12]

The form we normally associate with year of birth is a 4-digit decimal number, the first digit of which usually is a 1. If these 4 digits were held internally in a 6-bit binary coded format, then 24 bit positions would be required to store the year of the author's birth. Likewise, 24 bit positions would be required to store the year of his expiration, or 48 bit positions would be required for the

[11] Packing is the process of combining short data fields into one computer word so as to use the complete computer word most efficiently.

[12] Before going into this subject in further detail, the distinction between the *content* of a field and the *visual form* usually ascribed to that field should be kept in mind.

total. If the years were converted to 4-bit code, then 16 bits would be required for each of the two fields and 32 bits for the two dates, a saving of 16 bits or 33 percent!

If, instead of any binary coded notation, the year of the author's birth were converted to pure binary, then 11 binary bits would be sufficient to hold a date less than 2,048. The two fields could be held in 22 bits, a further saving of 10 bit positions.

If, in addition to the foregoing, some base year were specified arbitrarily, the year of the author's birth could be considered as an increment added to this base year. If the year 1000 were chosen as the base year, then any author born after the year 1000 could have his birth date expressed in terms of an increment to be added to 1000, such that the resulting sum would be the year of his birth. No man to date would have an increment greater than 963; therefore, 10 bits would suffice to hold the increment. In addition, since authors seldom live more than 100 years, the second field can be redefined as age-at-death. Seven bits would suffice for holding it, or a total of 17 bits would be required.

Thus we see that if we held 2 full decimal dates in 6-bit BCD (binary coded decimal), 48 bits would be required to store the information about the year of an author's birth and the year of his death. By adopting a suitable convention (the base year 1000) and an appropriate pair of definitions—the first field contains an increment such that the year of birth is obtained by adding the first field to the base year, and the second field contains age at death—we can reduce the number of binary positions required from 48 to 17. A similar phenomenon can apply in the case of alphabetic information.

It must be clearly understood that what we are discussing is how the information is held in storage private to the computer. Whenever information is required on a printout, on a reconstituted catalog card, or on a console display, it will be displayed as it now appears on the catalog card: a 4-digit decimal number for both the year of the author's birth and his death.

Nondeteriorating Files

If, when working with hard-copy files, such as a card catalog, an error is made in reading from the card, the chances of the same error being made in subsequent references are quite small. In working with a magnetic-tape file a different phenomenon is present. Magnetic tape files wear slightly as they become used. Although a piece of magnetic tape is good for many passes (in the thousands), there is a possibility that in some instance it will be improperly read. If this occurs when an update operation is being performed and a new file is being created from current information, then the new file will be written in error and all subsequent files will contain that same error. In short, a magnetic file deteriorates with usage.

This is a limitation of magnetic media. Each file is cumulative on the basis of all that has gone before, and degeneration is possible. But, as in many instances, recognizing the fault is half the battle. Current state-of-the-art magnetic-tape devices have built into them a series of checking circuits that guard against improper reads. A magnetic tape is checked as it is written, and if the write operation is not correct an error is signaled. These built-in circuitry checks are sufficient for most instances where a recirculating self-purging is involved.

On the other hand, since a reference file does not cleanse itself, reference files usually warrant special handling. The programming profession accomplishes this through programs which provide checks in addition to the hardware checks already available. These are called by various names, such as check-sum, hash total, or Orthocount. They are, in every case, techniques whereby the purity of the file can be guarded through the use of a little extra machine time and a little extra storage space.

A hash total works in the following way. Most of our larger computers can consider alphabetic information as data. These data are added up, just as if they were numeric information, and a meaningless total produced. Since the high-speed electronics are very reliable, they should produce the same meaningless number every time the same data fields are summed. The transfer of information within the computer and to and from the various input/output units can be checked by recomputing this sum after every transmission and checking against the previous total.

Some computers have special instructions built into them to facilitate this check, whereas others accomplish it through programming. The file designer considers the hash totals as a form of built-in audit. Whenever the file is updated, the hash totals are also updated. Whenever a tape is read, the totals are reconstituted as an error check. Whenever an error is found, the operation is repeated to determine if a random error has occurred. If the information is erroneous, an alarm is sounded and machine repair is scheduled. If information has been actually lost, then human assistance is usually required to reconstitute the file to its correct content. Through a combination of hardware and programming the validity of large reference files can be maintained even though the file is subject to repeated usage.

Summary Remarks on File Loading and Searching

Files are the heart of any library system. The shelflist is an inventory. The charging operation is a scheme for controlling that inventory. The author and title files are merely indexes kept in a specific order to facilitate selections. The subject file is another index kept to allow expedited searching. All of these combined form, in effect, one huge master file, which has been split by considering the requirements for the job and how best it can be organized for the tools and facilities available.

Any new system will similarly be designed for the tools and equipment available. Although the analysis techniques are the same, the end product will, in fact, be quite different. But, there will still be files. The files will be organized in some "efficient" manner. Efficiency will be required since the cost of storage for the files will be a significant part of the total system cost. Efficiency can be defined only after careful consideration of the requirements and the equipment available for the task.

If conversion of library files is undertaken in the near future, both magnetic disk and magnetic tape will be used for file storage. The master files will be kept on magnetic tape. The active files will be kept on both magnetic tape and disk. For operational information that will tolerate the delays associated with batching, magnetic tape will be used for its economy and efficiency. For information subject to selection, magnetic disk will be used. For information subject to searching, either tape or disk will be used, depending on whether batching can be tolerated.

For information subject to searching, the actual search will probably involve magnetic disk, since the search may be initiated in the middle of the file. The disk may be loaded just for this search purpose from magnetic tape. Thus, the disk may not be permanently dedicated to a single use; this again will be decided primarily on whether the delays of batching can be tolerated. In either event, the search speed, i.e. cost, efficiency, and throughput, is a linear function of the length of the average record for any particular hardware configuration. One important method for reducing the record length is the split-file concept (as noted, this has further advantages if the same information is held in multiple orders). Additional efficiency may be obtained by packing the file as densely as possible. This will conserve space and reduce the average entry length still further Many computers can be programmed to pack the requests and search the packed file, i.e. only unpack an entry after it is judged a "hit."

An attempt has been made to point out the costs and trauma involved in the initial loading of such a master catalog file. The existing file must be mechanically converted to a machine-readable form. This conversion alone will cost from 6 to 30 cents per catalog card. The file must then be edited and corrected, a procedure accomplished partially by machine; however, the cost of preparing the manual corrections will be significant. A change or update procedure must be established so that the file is not obsolete after the conversion is complete. The uses of the file will determine how it is finally structured, split, and packed. After the file design is complete, audit trails and check sums can be added to stanch any deterioration that might occur through extended usage.

A Plan of Action for Librarians

In closing it seems appropriate to suggest how to proceed.

1. Immediately adopt one of the existing techniques and automate the catalog cards at the source. If this were done in conjunction with automatic typesetting, an economy over present operations might result. In either event, the size of the file to be con-

verted would not be increasing and the mutual learning that must precede any large automation endeavor could be started.

2. A study should be undertaken to determine how accurate index files must be. First, develop some definitions; then develop a threshold measure of error by field to determine the minimum acceptable quality; and finally, establish a trade-off function of quality vs. cost, so that we may see what purity costs and judge how much of our limited resources should be placed here.
3. After the definitions noted in the preceding paragraphs are available, a statistically significant sample of the existing file should be taken, converted, and cleaned up. Meticulous records should be kept so that the error content of the entire file, e.g. the National Union Catalog, can be estimated. As this is being performed, cost records should be kept so that benchmark costs for keypunching (the method probably used for the sample) are available as a byproduct.
4. The library community should be stimulated to debate, in publication and open forum, the requirements for the master catalog file. In particular, the community should be encouraged to discuss, and eventually agree on, the following: "Resolv the present catalog card contains inforn tion deemed unnecessary in an automat system for reasons of economy. These ite are. . . ." Some of the topics to be d cussed are the myriad type sizes, fonts, a faces that have been used on catalog carc
5. Having benchmark cost data available, r quirements clearly in mind, and a measu of file purity, then overtures should be mac to manufacturers of scanning equipment t obtain an estimate of the portion of the de velopment costs for special multifont scai ners that the library community will be ex pected to bear (if any).
6. Not until the above steps have been com pleted can costs, budgets, and schedules b intelligently discussed.

First, we require a measure of the task to b accomplished. Then, we need to assay the tools available. If additional tools are required, their development cost must be determined, and the library community can be expected to bear the portion (if any) that is directly attributable to any requirements unique to them. Then, given a definition of the task and the tools, we may speak of budgets, schedules, and contingencies.

Reference

1. Bemer, R. W. Survey of coded character representation. Communications of the ACM, v. 3, Dec. 1960: 639–642.
2. A bibliography of glossaries. Datamation, v. 8, Dec. 1962: 19.
3. Bourne, C. P. Bibliography on the mechanization of information retrieval. Supplement 4. Menlo Park, Calif., Stanford Research Institute, 1962. 24 p.
4. Computing reviews. v. 1+ Jan. 1960+ New York, Association for Computing Machinery. Bimonthly. See especially the section entitled "Information Storage and Retrieval."
5. Durkin, Robert E., *and* Herbert S. White. Simultaneous preparation of library catalogs for manual and machine application. Special libraries, v. 52, May–June 1961: 231–237.
6. Isotta, N. E. C. A suggested first step towards automation. Aslib proceedings, v. 14, Oct. 1962: 333–341.
7. King, Gilbert W. Instrumentation and the information systems revolution. Computers and automation, v. 12, Feb. 1963: 22–24.
8. Vickery, B. C. On retrieval system theory. London, Butterworths. 1961. 159 p.

APPENDIX

A File Organization to Facilitate the Searching of Index Files

The Concept of Index Files

Index files are ordered collections of entries that describe a store of information. It is well known that the cost of disk storage for an information retrieval system is one of the largest component costs involved. Various schemes have been offered to reduce the storage space required to store a file of given size. Such methods involve reducing the length of the average entry.

Entries are made up of keys and respondents. The key is the set of fields on which requests will be honored. The responding field(s) will either supply the information to answer a search question, or will supply the location where such information can be found. Index files are ordered on the fields of the key to simplify both reference and maintenance.

With the above definitions, we can use the instance of a library as a clarifying example. The materials on the shelves of the library constitute the store of information; the card catalog is the index file. Two types of requests are honored: *selections* and *searches*. Index files are kept in some order or sequence. If a requester supplies unambiguous precise information for the fields on which the file is ordered, a selection may be made. For example, if the card file is kept in alphabetical order on main entry (e.g. author or title) and if a requester supplies the exact title, a direct selection may be made and the index information (the whereabouts of the book) obtained.

If the history of retrieval requests shows a strong statistical bias toward one or two key fields, then the file is usually kept on both of these fields to facilitate selections. For example, in the library the card catalog (index file) is usually maintained alphabetically by both author and subject to allow direct selections on both of these two categories. In an automated system, the above functions would be served by maintaining a master file (shelflist) on a nonserial storage medium (disk). The author subfile would be composed of a simple entry which would contain the author's name and the location of the full entry in the master file. The subfile would be kept in alphabetical sequence on the author's name. A request would be processed by selecting from the subfile the location numbers associated with any author whose name matched the name given in the request. The master file would then be referenced on the location numbers found, and the complete entries obtained. These would be "near hits." The near hits would then be processed against any other criteria supplied with the request and the "hits" output as the response.

In a similar manner, a title entry would be kept which consisted of only the title and the location in the master file where the complete entry could be found. The near hits would be selected, any additional criteria applied, and the hits output as a request response.

The split-file technique outlined below can be shown to involve the minimum storage capacity and also to offer extremely high-speed response to a retrieval request.

Other techniques are required when the requests do not statistically segregate themselves into a few popular classes. If the file is indexed in depth, then the file would need to be retained in many orders if only direct selections were to be allowed. At some point this becomes uneconomical, and a second type of transaction, called a search, is required.

Search File Criteria

The design criteria for a search file are twofold. First, the total storage is to be minimized. This calls for the average entry length to be minimized. With a search file this begets sophisticated programming techniques which utilize variable-length fields and entries. In addition, the fields themselves are encoded and packed to evenly load the computer main frame and the input-output units. Clear-cut criteria have been developed that define efficiency-of-search in terms of average entry length and component utilization.

A second important criterion for search files is the file organization. File organization is a generic term that depicts how each field within a file relates to other fields within the file and what that relationship is. Competing file organizations are judged on the basis of minimizing search time while holding storage volume constant. The technique outlined below shows how to minimize search time. It also reduces storage by use of extensive packing and variable-length handling.

First, it must be observed that the file must be cataloged by a competent person or an adequate machine process. The result of this operaton will be a series of descriptors (keywords, added entries). The descriptors will be encoded into some dense numeric character set (probably binary numbers). The length of the binary field will be fixed. It will be set to the next power of 2 greater than the number of descriptors in the thesaurus (i.e. authority file) the cataloger uses during the cataloging operation.

For example, a thesaurus might contain 10,000 terms. A fixed binary field of 14 bits would suffice to encode this thesaurus. The encoding would be applied to the entire file as it was constructed. When searching, requests would be similarly encoded prior to the search. (Note that the total volume of bits of a coordinate file is significantly less than the volume of bits required to store an inverted file. This is simply because the length of the field required to hold the document number is greater than the length required to hold the descriptor.)

Definitions

Definitions of the factors used are given in the following list.

1. Let the letter D stand for the Document reference number in the master index file.
2. Let the letter C stand for the descriptor Code associated with either a document or a request. If more than one code exists, let these be designated by subscripts, i.e.

   $C_1, C_2, C_3, \ldots C_n$

   where n stands for the number of codes associated with a specific document or request. The number of codes is not arbitrarily limited by the following procedure, but is left to the discretion of the reference analyst.
3. Let $\ell(D)$ be the length of the D field in bits.
4. Let $\ell(C)$ be the length of the C field in bits.
5. Let N_D be the number of documents in the collection.
6. Let N_C be the number of descriptors in the thesaurus.
7. Let $\bar{N}_C$ be the average number of descriptors per document.

The Coordinate vs. the Inverted Index

A practical case gives $\ell(D) > \ell(C)$, which states that there are more documents in the collection than descriptors in the thesaurus. Also $\bar{N}_C$ is greater than one.

If an entry consisted of a key and a reference to the master index file, then the average length of a coordinate entry in bits would be:

$$\ell(D) + \bar{N}_C(\ell(C)).$$

The total number of bits for the entire index would be:

$$N_D[\ell(D) + \bar{N}_C(\ell(C))].$$

Similarly the number of bits in an entry of an inverted file would be:

$$\ell(C) + \frac{(\bar{N}_C)(N_D)}{N_C}(\ell(D)).$$

The total number of bits in an inverted index would be:

$$N_C\left[\ell(C) + \frac{(\bar{N}_C)(N_D)}{N_C}(\ell(D))\right].$$

Thus an inverted file is always larger than a coordinate file. The additional bits required are given by:

$$N_D\bar{N}_C(\ell(D) - \ell(C)) + N_C\ell(C) - N_D\ell(D).$$

For example, a collection of 10 million documents and 10,000 descriptors whose depth of indexing averaged 8 descriptors per document would have:

1. $\ell(D)=24$ bits
2. $\ell(C)=14$ bits
3. $N_D = 10\times10^6$ documents
4. $N_C = 10{,}000$ descriptors
5. $\bar{N}_C = 8$ descriptors/document (average)

The volume of a coordinate index file would be: 1.36×10^9. The volume of an inverted index file would be: 1.92×10^9. The difference would be: $.56\times10^9$.

Thus a coordinate file has smaller volume than an inverted file. If the coordinate file is organized as outlined below, the search criteria are well formed. The irrelevant material may be easily skipped, so that only the relevant material is searched. Also, a clear-cut criterion exists for terminating the search, that bypasses subsequent irrelevant material.

Assumptions

The following assumptions are made.

1. It is assumed that the file is compressed through the elimination of nonsignificant zeros.
2. It is assumed that all entries are allowed to be variable length through the use of a word count associated with the entry header.
3. It is assumed that a simple, single level of search is required and that all of the descriptor codes associated with a document are weighted equally as to importance. (A following section will show the extensions required to release these two restrictions.)

The Search Index File

It is proposed that the search index file be kept in a special way. The *contents* of each variable-length record will be D, the related $C_1, C_2, C_3, \ldots C_n$, plus control fields as required. It should be noted that there will be only one record per document (e.g. coordinate entry file). This record will contain only pertinent information about a document: master file reference number and descriptors. Thus, only D and C_1 through C_n are kept.

Within a record, the C_1 through C_n terms will be kept as fixed-length data elements in *ascending sequence*, low to high, i.e.

$$C_1 < C_2 < C_3 \ldots < C_n$$

(Note that equal codes have no meaning.)

The records within the file will be *sequenced* on the string of C's considering them as a single variable-length key. The key is considered a *left* justified number. Where blanks exist, these must sort low to numbers.

An example: if D number 4002 had the following descriptor codes associated with it: 567, 234, 123, 345, it would have the following format before it was posted to the file.

D	C_1	C_2	C_3	C_4
4002	123	234	345	567

After the above document was posted to the file, the file would look like:

D	C_1	C_2	C_3	C_4
1000	123			
9000	123	234		
7000	123	234	345	
4000	123	234	345	567
4001	123	234	345	567
4002	123	234	345	567
3053	742	999		
0123	846	978	1235	
8421	847	1341		
9766	954			

As can be seen above, the codes are ordered low to high within a variable-length key. The records of the file are sequenced on D (the document reference number) within the key.

The Request

The request will enter the computer and be reformatted. The descriptors will be encoded into a dense binary set. Then, the descriptors associated with a request will be ordered low to high and left justified. For example, if the 321st requester requested all documents that had the following three codes: 978, 846, 1236, it would be stored in the following format:

RID Request number	C_1	C_2	C_3
321	846	978	1236

The formats, of course, would be variable length also.

The Search

Note that (a) batch size is limited only by primary memory available; (b) whenever a request is completely processed, it will be eliminated from further consideration, thereby speeding up the remaining processing; (c) when each retrieval transaction is completed, no repeated handling is required; and (d) after the search, the D numbers for the near hits are used to select the full index entry from the master file. If a sophisticated processing routine is used, a figure of merit can accompany the printout of hits.

Each request is formatted as depicted in the previous section. Search is initiated by locating the first entry in an area of interest. The monotone ordering of the codes within the key makes this possible. (Monotone is used here to mean that the series of code numbers is irreversible. Thus each descriptor code number can be equal to, or greater than, the code number immediately preceding but not less than it.) When the area of interest is located, the search is started in earnest.

Each request is compared against the entry from the file in the following manner. The entry is read into working storage. If any code from the request falls within the range C_1 to C_n of the entry, a compare subroutine is entered. The subroutine handles the detail comparing, the level associations, and computes a figure of merit. If the figure of merit is above some arbitrary threshold, the D and key (containing the descriptor codes) are stored as a near hit. An entry is held in memory until all of the requests in the batch have been processed against it, then the next cylinder of information is obtained from the disk.

If the file contains 10 million documents and if buffered operation is assumed at 900,000 bits/sec./channel, the process should proceed at full read speed until a hit is found. With four channels, a search of a tenth of the file would take 38 seconds. A batch of approximately 100 could be allowed without increasing the search time.

The scan subroutines will contain a test to eliminate each request from the batch when it has been completely processed. This will decrease the load on the computer and assure that the process, in the end at least, is limited by input rate. When a request is eliminated, the request and certain statistics concerning the hits will be retained for output. These statistics will be used to edit the volume of output and for management reports.

After all requests have been eliminated from the batch (the index file has not necessarily been completely passed), the hits are sorted. The new sequence is reference number within figure of merit within the request number. If the request itself has been awarded a fictitious D number of zero, then the final order is the request followed by the hits. After the sort, the master index file is entered on reference number, further processing is performed on near hits, and, finally, the hits are output.

Possible Extensions of the Search Technique

If differential weights are assigned to the descriptors so that the encoding carries the relative importance of each descriptor code, this information could be carried into the file by appending a weight factor element to each descriptor element. These would always occur in related pairs. During the ordering of the codes (such that $C_1 < C_2 < C_3 \ldots < C_n$), the weights would be moved also. Thus, if document 213240 discussed nozzles (1735361) for missiles (1142716), the entry would appear as:

AD	C_1	W_1	C_2	W_2
213240	1142716	1	1735361	2

In a similar manner if a requester wished to specify that he was interested in a document described by a complex form of terms connected by AND and OR, two control fields per term would be required. A relevance subroutine would use these weights in determining the figure of merit.

Thus, much of the editing function of the reference analyst can be delegated to the machine. An algorithm for pertinence can be devised so that only pertinent documents become near hits. If weights are awarded to the codes to retain meaning, then a sophisticated algorithm can be devised so that the printed listing of hits contains only topical documents.

735-898 O--64----5

CONFERENCE SESSION II

Libraries and the "Uppercase Limitation"

VERNER W. CLAPP
Council on Library Resources, Inc.

I want to take advantage of my position as the chairman for the day to say a word about the interests of the Council on Library Resources in this meeting and what may come out of it. We are all aware of the rapidly developing interest and activity in automation and mechanization in libraries. The Council has supported the study now underway at the Library of Congress. The Navy Pier study has passed its first phase, and a publication has been issued, which I am sure you will find valuable. Mel Voigt at the University of California, San Diego, has gone through one phase of his serial record operation and is going into his second. All around us there are evidences of people who are either doing or who want to do things in this area. Some perhaps are not quite sure of the best way to proceed.

This conference was suggested to provide an opportunity to consider developments in this area. One of the principal benefits anticipated from this meeting was the preparation of the state-of-the-art papers. So far, I am glad to say, everything looks as well as, or better than, was predicted. The state-of-the-art papers are on a high order of excellence and provide a basis for profitable discussion of next steps. I look forward to a stimulating discussion in the next few days among librarians who have actually had their hands in mechanization operations, and librarians who have not yet plunged in but who are anxious to plunge in, and technical people who will guide us to the facts. I am sure that these discussions will be effective in defining profitable areas for future consideration, as well as in educating us all generally as to the criteria that must be applied to any operation of this kind.

I am one of those who, back in the thirties, looked at EAM equipment and fancied that it certainly ought to be put to library work. We were fascinated by the great speed of that equipment at that time; this was its great attractiveness to us. How little we knew of speed in those good old days! What impressed us was the ability to do clerical operations such as sorting and printing. What bothered us, however, was the font of type which was available. So we used to travel up to New York every so often, and visit Thomas B. Watson in his office, and we would say to him, "Mr. Watson, we know you're interested in library work, we know what you've been doing out there at Montclair, New Jersey. Won't you please make us a machine that will print upper and lowercase? Then we can really do some useful things in bibliography." Then Mr. Watson would lean back in his chair and look benevolent, as indeed he was, and he would say, "Well, you boys know that I'm interested in library work, I'll see what we can do." Then we'd go away feeling warm around the cockles of our hearts, and think we would get a printer in the next couple of weeks that would print upper and lowercase. How naive we were! We did not realize that it would have cost International Business Machines a couple of million dollars just to develop this one machine for us. We thought it was just a matter of putting on a little longer type bar with a few more characters on it.

Well, this is "printout," and I ought not to be talking about printout at this session; I mentioned it partly because this is the only occasion I will ever have to mention it, but also to point out the close relation between printout and file conversion and storage. As long as Mr. Watson would only give us a character font in capital letters there could not be any great fervor to load. If I may say so, the whole picture of automation in libraries from the thirties right down to the present date has been controlled by that uppercase limitation. The reason library files do not exist in machine-readable form is that nobody wants to go to the expense of converting files when the output can only be printed in capital letters without even decent punctuation. We stand today at the point at which this whole situation may be completely changed.

At this point I want to turn the meeting over to your discussion leader and to the authors of the working papers. Unfortunately, one of the authors of the working paper, Donald Black, could not be here.

File Conversion: Prefatory Comments

I. ALBERT WARHEIT

International Business Machines Corp.

IR vs. Processing Applications

I am not going to repeat the technical material in the paper by Black and Patrick but I am going to try to elicit from you some discussion about the librarian's problem in getting started in this area. We should perhaps consider this question first: Do we want automation or not? One could rush in and automate everything, including the things that do not need automation.

Almost all of the conference papers emphasize the information retrieval aspect of automation. However, in the academic libraries, the libraries that many of you represent, I rather think that you are not really hurting very much in the information retrieval area. Of course in special libraries, where the researcher may need highly technical material, information retrieval techniques may be applied profitably. But I gather from talking to many librarians that processing problems and techniques are the primary concern. In a way, these processing operations offer the most promise for automation and present fewer problems and more immediate payouts than a very elaborate retrieval technique. If you begin with processing then, after having learned the methods of doing things and the capabilities of the equipment, you will be better able to tackle the retrieval problems. Furthermore, information retrieval, I think, is concerned more with the question of indexing rather than with machines, and I think the indexing question will be the one which will cause the greater difficulty.

Two Approaches to Mechanization

There is a certain basic fear about undertaking mechanization because it would involve, in some areas, a radical change in the way things are done. Librarians have been told not to go at this thing piecemeal. They have been urged to think of the consequences of each step, and to plan in terms of a total system. Certainly an organization like the Library of Congress, where changes might affect library operations throughout the world, must weigh many factors before undertaking any basic changes. We know of the long struggles that go on over the slightest change in descriptive cataloging. On the other hand, I have seen another organization with a very difficult retrieval problem literally frozen in fear for a period of 10 years; it has studied and restudied the problem but has never been able to start on a new approach. It is sometimes quite impossible to jump from a system that is a hundred years old into the jet age without going through some evolutionary process. If you try to make this hurdle in one leap, so to speak, you will never take the chance. We have to think in terms of priorities and in terms of steps.

The problem of trying to work out the loading and conversion of library files requires the determination of desired goals and the steps that have to be taken to reach those goals. There are at least two basic approaches. We can start with what I call the "special aspects" in a library, rather than the essential bibliographic record. There are many operations in a library which could be mechanized. Verner Clapp mentioned Mel Voigt's serial project—this is a very good example. There are basic library tools, for example, subject heading lists, classification tables, and serial lists which could be published, collated, and updated automatically. There is the activity concerned with the publication and distribution of catalog cards from the Library of Congress. These are operations which do not directly affect our ultimate retrieval problem which involves, of course, the total bibliographic record. There is a great deal that can be done in these smaller areas.

There may be a lot said about printing at this meeting, and I do not want to get too much involved, but I must mention in passing that many small and special libraries are getting started by using the computer as a printing device for producing their catalog cards. If this is done, bibliographic information is being captured for free. If a library can produce its catalog cards and at the same time get a machine-readable record, then it is getting the machineable record literally for nothing. There are other benefits in this printing approach. I have seen operations where the card is not only produced by computer, but the running heads are also put on in order; that is, the tracings are entered. Then the cards are sorted in sequence so that the filer does not have to go through a sorting operation before the cards can be filed into the catalog.

What I want to emphasize is that the initial input problem can be tackled, and, hopefully, if the system is properly designed, mechanization can actually be cheaper than present methods. Black and Patrick point out again and again in their paper that the capturing of the bibliographic record is the easiest thing to do; it may even save money, but more important, it will produce the record that you must have later on. The problem will not just be one of hardware; the librarian will have to ask the question: Are there sufficient records in the file to enable me to make use of a computer? Even if you have the best system in the world, if you delay loading, it will take a long time before there will be a sufficient amount of machine-readable material for your system to be worthwhile. I believe that Patrick indicated in the paper that there is a 12 percent growth in the National Union Catalog per year; this adds up to a tremendous amount in a few years' time. The longer librarians delay in making this initial decision, the more difficult the task will be. The small start will also give one the benefit of experience.

The "Large File" Problem

Today most of the work in information retrieval has been done with the small file—10,000, 100,000, 500,000 entries, and so on. Some of the systems that are being designed today for the large library are mere extrapolations of these small systems, and quite frankly (at least I feel this very strongly) they are going to become uneconomical. Pumping all this material through is going to slow down the operation to the point where people will not put up with it.

People are just now beginning to think very seriously about how to approach the very large file. Again, Black and Patrick have covered aspects of this. It is not just a library problem; there are many large files both in government and in business. We are rapidly developing a lot of experience with the 10-million and the 50-million record file, and this is a problem which will probably be solved during the next few years. The hardware for it is being developed; the hardware available today is not really adequate. Patrick says we need about 14 billion bits for the National Union Catalog. That is about 2⅓ billion characters, and our present random access files are about 56 million characters. Nevertheless, I feel that the hardware will be here long before librarians are ready to use it.

The question was raised last night about the queuing problem, but I am less concerned with this. I look at an airline reservation system that has 1,200 consoles with a 2-minute response time. A lot of money is spent for this, and I do not know whether libraries can afford it or not, but the queuing problem will probably be solvable, at least if you have the money for it. The programs for handling multiple access are being developed.

In other words, the hardware, the software, and, as someone once said, the "crunchie-ware," will be here long before the files are put together. The real question is: When are we going to start and what are we going to do? This does not mean that you are not going to question the hardware capabilities. This is very important, and I think pressure on such areas as output printing and programming languages is very necessary.

Aspects of File Conversion

The capturing of new information in machine-readable form, then, is very promising and librarians can move ahead here with a fair degree of confidence. As to the conversion of older material, I am not nearly as sanguine about it as the authors have been. They may have used the National Union Catalog just by way of illustration, but when I looked at that catalog with respect to converting by photoscanning the cards, I had the reverse impression. I estimated that 10 percent might be scanned, and 90 percent would have to be keypunched. I was concerned not only with the type fonts, but also with the clarity of the printing on the card—the broken character, the smudge, the legibility of the handwriting, etc. Having seen present-day optical scanning operations and realizing the controls that have to be exercised to get error-free output, I am not really quite as hopeful as some of the optical scanning enthusiasts might be. This is a personal observation, but I do think the problems encountered in conversion have to be considered.

I started out by saying that I felt the academic libraries were not really hurting so much on the retrieval side. Actually, if you were to take trays of catalog cards, put them into the computer, and use the machine as a reading and printing device, I do not believe the results would be worth much. Unless you use the manipulative power of the computer by deeper indexing, you will not retrieve much more than you would by going to the guide card and finding the entry directly under the subject heading. For retrieval purposes, existing tools will do the job in many respects as efficiently as a computer, except for the output printing. It is only when deeper indexing and manipulative techniques are applied that more will be extracted.

I am not frightened by the conversion problem, because I do not know how much you librarians will want to convert. This is a question that should, however, be considered from the librarian's point of view; you should know first what you want to convert and how far you want to go in converting. The science librarian has an easier problem, since his material has a half-life of 5, 10, or 20 years, depending on the discipline in which he is working; in many areas he can ignore the conversion problem. The academic research library cannot.

I want you to be aware that capturing the bibliographic information is not just for the purpose of alleviating in-house processing work; it will also provide outputs that can be used in many ways. Consider, for example, that the printing of book catalogs, the announcement of new serial titles, and so on, will be greatly speeded up if one starts with current materials.

The meeting is now open for discussion.

General Discussion

HEILIGER: We had General Electric help us in the first phase of our study, and they analyzed our costs in great detail. We were alarmed at the size of our filing costs and at the cost of the proliferation of departmental libraries. As a result of that we became convinced that we should have completely centralized library services with printed catalogs that could be distributed widely around the campus. We think that the elimination of both the filing costs and the duplication costs in setting up branch libraries could more than justify this computer-based system.

WARHEIT: Yes, this approach to mechanization has made some librarians honest for the first time in the area of economics. At the Atomic Energy Commission Library I microfilmed material and gave it away rather than circulate it. Several librarians spoke in horror of spending 25 to 50 cents to microfilm and reproduce material to throw away. So I asked them what their circulation costs were, and so often I was told that they were practically nothing, but when you really looked it was about a dollar. This can be an opportunity for you librarians to take a hard look at what you are paying now. What are your filing costs? How much time do your catalogers spend in getting up and walking to the file and sorting through the entries, recording data, then walking back to the desk and re-recording it and correlating it with the book at hand? What is all this costing? These are very real questions that we haven't faced up to; all we can see is that that computer costs so much rental per hour!

MINDER: I would like to ask the librarians what they do when they can't get LC cards? I understand they use an ordinary typewriter which I don't think has over 64 characters. We don't bother with Linotype machines; we accept what's available from the typewriters when we can't get LC cards, and there are no complaints at my institution.

I'd like to comment about the value of considering the LC proofsheets[13] as a starting point. The LC proof is a tool which we use in our cataloging. If it is complete, we accept it as it is; if we want to modify it, we do so. But it is a temporary tool. If LC proof were available in machine-readable form, we could use it as received or improve it as time goes on without any permanent harm to the system.

TAUBE: At meetings like this we invariably get a standoff between the machine man who says: "You tell me what you want, and I will do it," and the librarian who says, "Tell me what your machines can do, and I will see if they fit." Now Patrick has said that he believes he could program the rules for filing or the rules for descriptive cataloging in a dictionary catalog. Now he can't do the latter because there is no agreement on what the latter are. In other words, the library profession, since the appearance of the ALA revision, has not agreed on its cataloging rules. Now I would put this question to the librarians here. Let us suppose that there could be a machine program for any agreed-upon rules, regardless of their complexity. Would this influence the librarians after 15 years to come to some agreement as to what the descriptive cataloging rules should be?

ELLSWORTH: I would like to say that the reason we librarians have so much trouble about costs is of course partly our own lack of ability, but in part because the problem is truly an elusive one.

With respect to the problem we are talking about today, we librarians cannot always decide how much of the problem has anything to do with machines at all, and how much of it has to do with organization and use of talent. So instead of trying to tell you nonlibrary experts why we hurt, I think we should try harder to tell you *how* we hurt.

If we can find cataloging information for materials that we have acquired and if we could batch the two without spending a lot of time doing so, i.e. identifying the book we have, getting it together with the catalog information, et cetera, then I think most of us would feel tolerably comfortable about the costs involved in getting our catalog made from that point on.

But that isn't what hurts us now; what hurts us is that we are acquiring all kinds of materials that we don't have any way of identifying. It costs us a dollar or two even to write and find out whether the Library of Congress has a card for each item. We have difficulty because we don't know how to organize the personnel that catalogs material in unusual languages. This problem probably has nothing to do with hardware, and yet it is a problem that many of us are worried about. We are buying books in uncommon languages, from Indonesia and so on. We can't possibly assemble enough people either in our acquisitions or cataloging departments to handle materials printed in these languages. Nor have we yet found a way of solving this problem in a way that we really know to be sensible—namely by putting it on a fully centralized basis. We know that we ought to do this, but we haven't been able to figure out a way of doing it satisfactorily. From this point of view the problem would seem to be a governmental problem more than a machine problem. It wouldn't matter if the information were available in the National Union Catalog, either compiled by hand or by machine, because if we couldn't identify in our library the book that came from Japan or Pakistan, we would spend a lot of money trying to match these things up.

[13] The LC proofsheet is the final galley sheet run off just before the individual catalog cards are printed. There are generally five cards per sheet; the cost is 4 cents per sheet or $60 a year for all the proofsheets. The proofsheets are issued in very broad classes, e.g. technology, literature and language, etc.; entries on the sheets are random. Many research libraries subscribe to this service as a mean of keeping abreast with new publications.

Now there was a time when the Library of Congres was working on a project called cataloging-in-source, which if carried to its ultimate conclusion might have helped to solve this problem for us.

What I am trying to say is that we librarians are really not very much interested in how the Library of Congress handles this automation problem if, at the same time, it will solve the specific kind of problem that I have been talking about.

WARHEIT: Mechanization is going to help in this area because now you must have a single, essentially a main entry, identification. In the computer it makes no difference if you have multiple approaches, and this can help in identifying the material in hand, because you don't have to determine the main entry; there can be several approaches to enable one to find the material.

ELLSWORTH: If it is in a language that you don't know anything about, none of this does you any good.

WARHEIT: Actually, you do have certain information. It is true that sometimes a book does wander in with no record at all, but more often than not you ordered it. You started with some piece of information that can be latched on to. I don't think it is quite as terrible as you picture, but maybe I'm minimizing the problem.

NEAL: I know of a company that decides about every 4 years that they will get a computer. In order to convert their operations, they go through a complete systems study and make necessary revisions, perhaps in their management, or in their reporting structure, or in the processing of materials. When they do this, then they turn around and cancel the computer. I'm not really sure whether the library would need a computer or not.

WARHEIT: Sometimes the main benefit from trying to set up a computer system is the fact that you set up a *system* and you clean up a lot of your problems.

HOWE: At the risk of being simpleminded, I'd like to say that I'm not sure you need a computer. In my library we decided that we didn't need all the information that is usually on a catalog card; we are putting one line of information per title on a punched card. We have 25,000 titles that are actually operating under IBM circulation control now, and we have an additional 165,000 titles on IBM cards which we can convert to our circulation control. We didn't approach just one process; we integrated the library routines from an administrative point of view. We do our registration, statistics, circulation control, ordering, cost accounting, and all kinds of use studies with data processing equipment. This is just in a little town of 80,000 people, with one central library, two branches, and two bookmobiles.

At the end of our fiscal year I took the six major routines that we do and computed that they cost $1.50 per hour per procedure on the basis of an 8-hour day. By procedure I mean that registration costs us $1.50 per hour. And what system do we have? Just the series 50 with the good old 402, an 082 sorter, a collator, and two keypunchers. I'm not sure you need computers.

WARHEIT: I feel this limited approach is fine at certain levels of operation, but when we talk of a national problem and of the very large research library with collections in the millions, I don't think that we can afford this compromise. True, this compromise will do a tremendous amount for you and it is effective, cheap, and exceedingly useful. Would you care to make some comments about this, Dr. Richmond?

RICHMOND: I am using the one-line entry for about 23,000 books out of 50,000 as an adjunct to the main catalog. For what I'm using it for, which is to take the catalog to the professor's office, it is all right. But if I were using it for more cataloging production, the entries would simply not be full enough; they are abbreviated much too much. I am having terrible filing problems. I just can't get enough out of one line.

WARHEIT: In one of our facilities we have this one-line book catalog, and it's wonderful. Everyone has access to it, and it has opened up areas for library services that were not available before. But let's face it, it is a very definite compromise in terms of bibliographic control.

ANGELL: If I could have the privilege of post-editing these remarks I would venture to describe what we are doing when we construct a catalog card. We are writing a formalized text which is the description of a document. What we put on this card enables libraries to respond to two basic kinds of questions that are asked of the store. First, the reader wants to see a book whose existence is known to him and enough of its objectively determinable and recordable features so that he can specify it within a tolerable range of ambiguity. The second is the reader who wants a document that we have and that would be useful to him if he knew of its existence. For this we provide aggregations of bibliographical description.

Now whether this is done in the future on 3 by 5 cards, or paper stock, or with whatever kinds of marks, it seems to me that this function of the entry will have to be available to libraries. It seems to me that, as a generalization, the first response that we want to make to the second kind of question, the subject or category, is a display of descriptions. We do not want to display the text immediately. We don't want to be able to push a button and have all of the biographies of Napoleon come down the chute. We want to be able to display descriptions, so that th user can make a selection. He is the only one who can do so.

DUBESTER: I think it can be stated this way: primarily the catalog serves as a finding tool; this is true whether the catalog is in a card form or in a book form. The traditional dictionary catalog, on the descriptive cataloging side, brings the works of an author together and the editions of a work together. On the subject side it identifies works on a given subject and works related to that subject. Rules are developed because they serve these functions of the card file. When you have a book catalog in a fixed sequential array that cannot be modified except by a new edition, you do not really try to serve the function of identifying all the works of an author and all the forms of the work. Unless you have a book subject catalog which is in a highly cumulative series, you cannot list everything you have in the library on a particular subject and everything related to it. These two functions are ideally served at the moment, within the limitation of size and convenience, by the card catalog. The book catalog does not really achieve this in as efficient a manner.

CLAPP: Let me in my turn say what this little piece of cardboard is. This little piece of cardboard is *two* things in one and this is its great merit and achievement. This is what constitutes it as one of the prize bibliographical inventions of all times. This card places a book, a bibliographical item, in a specific place among *all* the other bibliographical items in the world, so that you can find this item among all the others, all the millions, if you simply know the rules by which this card was constructed. It does this not only in one series but in a number of series: it does it by author; it does it by subject, usually several subjects; it does it by title, by a formalized title (now so formalized you can't recognize the title page from it, unfortunately, but that doesn't upset the principle); and, finally, it sets the book on the shelf in a classified order, among all the other books that have ever been or ever will be printed. This is a very fine achievement for that little wretched pasteboard to do, but it goes further than this. By being on this 3 by 5 pasteboard, it can now find its way into the trays of every catalog of every library in the world which has adopted this standard, and this may be 95 percent of all the libraries in the world, and this is a pretty fine achievement. In these trays it will respond to the various questions which are likely to arise as to whether there is a bibliographical item, among all the others in the world, which responds to the following inquiries: Is there one by this author? Is there one on this subject? Is there one related in a hierarchical classification to others before and after?

BERUL: I would like to comment with respect to cost. As Patrick indicated, it might cost 3 cents to produce the card with present technology. This is really very small in terms of the real cost of the cataloging operation. The real cost is the intellectual cost, and this is being duplicated in many libraries throughout the country. I think that the concept of printing catalog cards and making them available to libraries throughout the country has been a great achievement; I agree wholeheartedly with Clapp. But let's look to see if there are limitations that automation might remove. As I look at the present card service I see two limitations. One is the response time in getting this information out to the library before it has already ordered the book or is doing the descriptive cataloging. The second is coverage; 40 to 60 percent is fairly good coverage, but if we could have some method of cooperative cataloging, this percentage would be increased and the response time decreased, because the first library to report with adequate descriptive cataloging would in effect say, "All right, cataloging has been done once; let's quickly get the result out to all the users and see that it isn't done again." This can be done very cheaply, but it is the intellectual costs that we should think about rather than the 3 cents for printing out the card.

McCARTHY: I'd like to add a third benefit which I think should come. This is the so called "write-up" of the extra cards, the multiple typing of subject headings and secondary headings. It seems to me that somehow we should get over this in the machine age, and yet, unless I'm mistaken, we are all doing it over and over again.

WARHEIT: I mentioned earlier a program which does this job. The printing of the tracing on the top of the card and the sorting into filing order sequence is a real cost to the librarian. There are a number of special libraries that have justified their total mechanization on savings in this one area.

LUNDY: May I contribute just a brief note of information on McCarthy's question as to why we can't get a machine to print in all the tracings and the headings on top of the unit cards. My neighbor, Ralph Parker in Missouri, is using a Flexowriter complex of three machines; he has succeeded in getting the Flexowriter to work under direction from two punched tapes at once. One tape has the text of the unit card which is produced by that machine and the other tape instructs the machine to print the headings. This is all done automatically; I have watched this machine at work. And so, prior to the installation of a computer, apparently Parker has solved our problem.

PATRICK: IBM has had a machine for at least 10 years that does this.

R. D. ROGERS: Since we're back on this subject, there are two or three things I'd like to add about the card service. It has been inferred that there are old LC cards that are not in print. This was true, but since we have started to use Ektalith, we are now able to supply cards for anything that we have in our master file. Secondly, you might be interested, incidentally, in the production cost per card; our Ektalith cards are costing around a half cent a card; the regular printed cards cost just over a cent a card, so it obviously is not the cost of a piece of paper that is making it expensive for you to get a set of cards.

Now with respect to the overprinting of headings: You know that we are trying to encourage wholesaling of catalog cards, and it is conceivable that if this were to go forward in a big enough way, it would be possible to supply sets of cards as H. W. Wilson does, with the headings overprinted. I think we would still have the problem, operating without the kinds of machines that are inferred here, of getting the overprinting for the odd set of cards that someone might want, say for a 1915 imprint.

BERUL: There is an answer to McCarthy's desire to minimize the manual retypings for tracings at the remote library. Warheit mentioned the 7090 program and the Itek Crossfiler. On my last trip to the LC processing section, I saw how they put the tracings on for their own cards and they do this with a Multigraph. Why aren't these sets made available by printing technology and production techniques for other libraries or is there no demand for this kind of service?

R. D. ROGERS: As you probably know, the H. W. Wilson Co. does sell cards with headings already superimposed on them; these are very popular in certain libraries. I think some research libraries do not necessarily want LC headings. This would be one problem. Also, to be perfectly frank, it is all we can do at the moment to keep up with the demand for LC cards. The increase is running 10 to 15 percent a year. We are now selling over 45 million cards a year. As someone has said, we are getting to the point where it is almost impossible for enough people to get their hands into the trays to carry on this operation. Lots of people who have looked at the Library of Congress Card Division have said that this is a logical place for automation. I don't think that this is really the sort of thing this conference is about, but nonetheless I think that nothing would be more important, in a pragmatic way, for the Library of Congress to do than this. It might be possible for us to produce cards with the headings on them if this is what libraries want. I don't know that I should go beyond that.

GULL: I'd like to address a question to Patrick. You have said in your paper that you have developed a unique file arrangement which is given in the appendix. I am surprised that no one has brought this up prior to this time. Would you care to support that further?

PATRICK: For the detailed material it would be best if I talked to you outside this meeting. I'll be glad even to flow chart it for you and show you the benefits in either mathematical form or conceptional form. The provision for search strategies must be built into your original files, into your original input of your file. You must decide what searches you will allow in order to have the input data available when you wish to perform the search. There are many schemes for this. We have found that present library files are organized so that the two kinds of searches that we like to do are convenient; the file is organized by author-title and by subject.

There are times, howev... ...hen you may want to go into indexing in depth and hierarchical forms of indexing. You will need different search formats from time to time, but these may not occur sufficiently frequently to have an entire file for each of the possible formats. Therefore, you are always faced, eventually, with some search strategy for which the file is not predominately organized; a scheme for this situation is described in the appendix to our paper. I can't guarantee that it is unique, I have never seen it in print before, and I developed it.

We have been talking, rather fashionably, and otherwise, about justifying a computer. If you have a large enough problem, that is beyond the 80-column card, you need the variable-length formats that modern computers supply. If you have a computer like this, it is going to cost from $200 to $600 per hour for every hour the power is up; whether you buy it, beg it, or steal it, it is costing somebody that kind of money. Consequently, you are concerned immediately with the efficient use of this equipment, just as if you were running a large manufacturing shop. The search strategy in the appendix allows you to utilize the balance of the computer so as to exploit your resource efficiently. I will go into this in more detail with smaller groups.

WARHEIT: It is true that we have to identify the elements that we want to search against; this is the input side. Having done that, however, the organization for efficient utilization of the equipment will vary somewhat, and there are, as Patrick has indicated, various patterns of efficient organization. I have a scheme, too, and there are others who have schemes for their operations. I do think that we should be concerned with the identification of the elements and how we arrange them, shuffle them, and set them out for efficient utilization when they are needed.

DUBESTER: I think it is relevant to relate what Patrick said to an earlier question, posed by Dix, about getting the correct entry for the item in hand. Suppose that the file is so organized that in the computer store there is an array of main entries or unit entries—in other words, author, title, imprint, collation, and so on. Given that situation, the necessary consequence is that somebody in Dix's library will have to prepare a similar entry from the book in hand and then search that store to find the item that matches it. Patrick suggests another possibility. You have this array in some order, but for every item you have a unique identification number, an addressable number. You have another file which contains just authors, and for each author you have addressable numbers. You have another file for titles, and another file for subjects. The person in Dix's library can say, "I have a book with this author and this title on the title page." The searcher does not go into the main file but into some subfile to seek the common identification number for the given author and title. These files are searched and matched, the main file is searched, and you come out with the main entry. You didn't give all the information that is in the main entry, but you asked for special searches to get certain combinations from what Patrick describes as compacted files. The search is not made in the whole file but rather in partial files which are compacted for efficient utilization of machine time.

This is the problem of file organization and file structure to make the most efficient search, and it does involve a prediction of search strategies in terms of the potential demand. The point, however, to be emphasized here is that there are a variety of files that can be generated when you have an automated store; these are not necessarily the files that you have with a conventional card catalog. Many of the rules that we have developed have been based on our work with the dictionary file. The type of catalog is a significant factor in the development of the rules that are going to be used by catalogers. You can include the dictionary catalog in the automated store, but you can also develop other files to serve different needs.

WARHEIT: I mentioned this morning that some large files are pure extrapolations of the small files. The more efficient files today are being organized so that there are a whole series of tracings with only their addresses and there is a separate total bibliographic file. You search whatever tracing you want to select, get the address or item number, and then get the total printout from the complete bibliographical file. Now once you have these individual tracings and addresses in this compacted file, then you can arrange and assemble the file in any way you want to suit the efficiency of the operation.

LOGSDON: Is there a possibility that persons working on the machine side can develop some kind of code, with relatively few arbitrary symbols, which could be applied with reasonable accuracy to any piece of paper, book, document, or mimeographed piece, and which would match centrally with whatever system of coding and arrangement was used? In other words, a hierarchy of 10 or 12 symbols that might discriminate one item from 10 million or 50 million?

WARHEIT: Unique identification is a very difficult problem. We have this, of course, in the identification of people; we have names, you know, but names aren't too good, and so Internal Revenue is now starting to use our social security numbers. In Sweden everyone is assigned a number at time of birth. Unique identification is a very difficult problem, and I don't know that the machine people would be the ones to answer that. Certainly if LC were setting up, for instance, a coding system for the LC classification, the machine people might try to persuade LC to stop using mixed notation. But I think the problem of unique identification is really a librarian's problem; the reduction of that to a code would follow. I agree that if a book came in with a machine-readable code on it which could be put under a reading device and automatically matched up with the LC card number, it would be fine. That is your cataloging-in-source.

ROSE: We have been discussing the National Union Catalog; we have been discussing the Library of Congress; we have been discussing computers; we have been discussing intergalactic communication of card catalogs. I think that, however trite it may be, we ought to keep in mind that there are intermediate steps that can be taken that do not necessarily involve computers, and that do not necessarily involve cooperation with the National Union Catalog. These intermediate steps might be better solutions for some of the smaller libraries or for some of the more specialized libraries.

PATRICK: Why aren't you librarians already doing it? That I don't understand.

WARHEIT: Many of the smaller libraries are starting in this area.

PATRICK: But the electronic accounting equipment has been in the field for 20 years.

WARHEIT: Yes, but librarians didn't know how to go about using it. I can speak for myself, because when I had EAM equipment and tried to work out problems the IBM salesman didn't tell me that they had a 101 machine. I, of course, had no way of knowing about it, and therefore couldn't and didn't apply it. In other words, I couldn't communicate with them. Not until the pharmaceutical industry began using the equipment were my eyes opened to some of the potentialities. I was being critical this morning because, quite frankly, when I tried to use some of this old EAM equipment it wouldn't work. I couldn't tell the machine people what I had to do, and they didn't tell me the capabilities of their machines. Now we have learned a great deal, and we are in the process of applying these techniques at a special, restricted level. But again you have the human inertia problem, and there is the serious problem of educating people about machine possibilities.

SPARKS: I think part of the problem is that as librarians we have not recognized the tools that we have for what they are. This is what has held up our use of the machine. We haven't been able to interpret our needs in the proper terms.

PATRICK: It looks as if we have been so busy doing the work that we haven't looked to see what we are doing. As was said last evening, if you can define it as a clerical or formal operation, then you can mechanize it. But you haven't *defined* it as a formal operation.

EDMUNDSON: As a member of the survey team studying the operations of the Library of Congress, I can tell you that we went through the routine that was alluded to earlier—where the librarians wanted to know what the machines would do and we wanted to know what the library problems really were. I would like to point out that the cycle is much more complex. The problems can be stated by librarians; the computer people can respond; it turns out that the proper response is not a single solution, but a set of alternatives. We then found that a cost analysis was missing; we had one made. The report of the survey will include some of the results of this cost study of certain Library of Congress operations. It is a very involved study, and I regard it as one of the most important pieces of work that was performed. The numbers reported may not apply to other libraries, but the cost methodology can be used to produce costs for individual library situations. I do believe that the ultimate decisions are not going to be made by librarians, nor by computer experts, nor by the cost people, but instead by the administrators of the funds, who will act upon the various alternatives in the light of the cost.

SWANSON: This is a slight *non sequitur* and I apologize for my delayed reaction. This is in reference to a slip of the tongue earlier this morning when Warheit said "millions" instead of "billions." It doesn't scan too well, but here it is anyway.

An IBM salesman named Ben
While pricing made a slip of the pen.
He sold library automation
For the whole bloomin' nation
By dropping 17 factors of 10.

WARHEIT: We have broad shoulders. One more point before we conclude—the real reason that the librarian hasn't defined his problem to the machine people is because he has not tried to put his operation on the machine. The minute he starts putting some specific operation on the machine, he starts defining his problem; and he defines it very well. You can sit back and theorize and try to define your problem, but you're not going to do it until you start getting your hands dirty.

SECTION III

File Storage and Access

Automated Storage and Access of Bibliographic Information for Libraries

RICHARD L. LIBBY
***Itek* Corp.**

Introduction

The application of technology to the storage and retrieval of information is an area that has received much academic, industrial, and governmental attention within the past two decades. The most rapid and successful exploitation of technology has been in the handling of information which is characterized by three attributes. First, it has been primarily quantitative information. Second, it has been information that could be segmented and labeled with reasonable assurance that the labeled segments would match its subsequent use. Third, the value, the time, and the frequency of use of the stored, labeled, information segments could be reasonably predicted.

The bibliographic material of libraries, the descriptive material for library holdings, falls within another class of information handling. It is information that is not primarily quantitative in nature. It is not easy to predict the value, nor the frequency and time of use. It is, however, susceptible to being segmented and labeled with reasonable assurance that the labeled segments will match subsequent use. But even this assumed susceptibility, which has similarity with information with which automation has had success, is suspect. It can be cogently argued that use of bibliographic material with subject, author, and title headings as labels is a marriage of necessity. Indeed, libraries deal with information that is not readily amenable to formalized prediction as to how it will be used, how often it will be used, and how users would like to ask about it. In this sense, automation of bibliographic information is in that class of information-handling problems that includes the storage and retrieval of management decision-making information, certain military command and control information, intelligence information, and so on. The common overall characteristic of such information is that it is expressed and described by the whole domain of human language (including numerics).

It is sensible to ask now: Will the techniques and the technology that have been so successfully applied in the past be suitable for automated handling of bibliographic information? It is also pertinent to ask: What techniques and technology in automated file storage and access appear best tailored for application to the automation of bibliographic information? It is the purpose of this paper to attempt an answer to such questions. To do so requires that the exposition range from the tutorial to the speculative with the attendant risk of causing boredom on the one hand and strong disagreement on the other. The latter is welcomed since selection of the proper course for automation of bibliographic information is most probably tantamount to selecting the proper course for a future generation of information processing.

The Measure of Information

Prior to consideration of either existing or future automated file storage and access methods and devices, it is worthwhile to discuss the terminology of the trade. First and foremost is the quantitative measure of information. Measures of information "value" have yet to be devised, but the work of Hartley and Shannon has provided a measure of the quantity of information. (For papers by Hartley and Shannon see items 5 and 12.)[14] Just as nature found it desirable to meas-

[14] This and similar references refer to items in the bibliography, page 88.

ure the intensity of a stimulus to our eyes or ears by producing a nerve message response that varies as the logarithm of the intensity of the stimulus, so it was found that a logarithmic relationship was useful in defining a unit quantity of information. It may be recalled by the reader who does not frequently use mathematics that the logarithm of a quantity to the base 2 ($\log_2 Q$) has a value equal to the number of times the integer "2" is used as a factor in multiplying by itself until the product equals the quantity (Q). Thus the logarithm to the base 2 of 4 is 2 since $4=2\times2$; of 16, four since $16=2\times2\times2\times2$ and so on. The information content of a symbol, a letter, or a word is simply equal to the logarithm to the base 2 of the number of equally possible choices one has in selecting (blindly) the symbol, letter, or word from all possible symbols, letters, or words. The unit of information is called a "bit." For example, a single isolated letter of the alphabet contains an information content equal to the logarithm to the base 2 of 26 ($\log_2 26$), where 26 is of course the number of equally possible choices available from the alphabet. The logarithm of 26 is 4.7, hence one letter (in isolation) has an information content of 4.7 bits.

If there is "noise" present, for example, the letter is smudged, or if there are constraints (say one had previously selected a *q* then only a *u* could occur next), then the average information per symbol drops. (See item 4.) Unlike the situation in communication channels, one can usually assume that in data processing equipment the occurrence of an event (or recorded mark, a voltage pulse, etc.) or the absence of occurrence of an event (a recorded mark, a second level of voltage, etc.) is detected with certainty. The machine expects either occurrence equally and hence in such a case one bit of information is equal to one binary digit.

The Machine-Readable Representation of Information

Since an alphabet character "contains" 4.7 bits of information for an "unexpecting" machine, then a sequence of 5 on-off or 2-state events (called binary digits) would represent the 26 letters of the alphabet plus 6 other symbols such as punctuation symbols and space ($2^5=32$). Indeed, early teletype systems used 5 positions (with a hole punched or not punched) across the width of a narrow paper tape to represent information. It is obvious that if upper and lowercase letters, numbers, and special symbols ($, ¢, @, etc.) are to be represented, then more than 5 "holes" or binary digits (2-state events) in a group (also called byte) are needed to have each group represent say 80 plus such symbols. In this case, the 5 "holes" would still be adequate provided the "shift key symbol" or special operating byte technique were used. Here, one of the 32 bytes is specifically prohibited from representing a character symbol and the equipment circuits are provided to "recognize" the occurrence of this byte and treat some or all of the subsequent bytes as new characters until the occurrence of the same or another special byte occurs. Alternatively, if one is sure that certain symbols will never occur together in a sequence of information being processed (say "*q q*" or "*z x*") then this sequence can be used for this same operational function.

Obviously these techniques require special circuits in input/output mechanisms or greater information storage space if they are frequently needed to represent the stored information. There is an increasing trend towards the use of 6, 7, and 8 binary digit bytes for handling alphabetical and numerical data (frequently called alphanumeric or alphameric). Although binary (1-state or 2-level) representations of information within data processors is most common, ternary (3-level) and other multiple level representations are possible. The binary representations within a machine can be 2 levels of voltage, current, degree or polarization of magnetization, opacity of photographic material, etc. Whatever the internal representation, the symbolic representation is usually by means of a "zero" (0) binary digit and a "one" (1) binary digit. Various codings of symbols, characters, and numbers are possible, using sequences of zeros and ones. A discussion of these possibilities is a topic in itself and outside the scope of this paper.[15] It should also be noted that other forms of information including voiced and imaged material are frequently expressed in digital form. The reader is referred to a companion paper, "Library Communications," by Emling, Harris, and McMains for further information on this point.

[15] See almost any general reference manual on available data processors.

Human Information Processing Rates

Inasmuch as automated information storage and access mechanisms ultimately must communicate their information to human beings, it is pertinent to consider the human information processing rates involved.

It is important in man-machine system design to engineer an appropriate match between equipment performance (e.g. system throughput rate) and human performance. In the case of human beings there appears to be a more fundamental limitation to the rate of conscious processing of information than that imposed by their input information channel capacity (e.g. the visual field). A number of experiments have been performed that demonstrate a reasonable upper limit to the human brain's conscious information processing rate of about 25 bits of information per second. (See item 7.) Some investigators say that this may be as high as 40 to 50 bits per second. (See item 11.) There is also experimental evidence that if the human being must perform associations between things (symbols, words, etc.) that human information processing rates approach an order of one bit per second. The reader is reminded that in the matter under discussion one bit of information is not directly translatable into one binary digit (as in the general computer case). For example, because of the many constraints inherent in language, of which the human being is well aware, the individual alphabetical symbols in a running text may convey to a reader an average of a little over one bit of information, whereas a computer circuit requires five binary digits to recognize the symbol.

The Technology of File Storage and Access

Most data processors currently in use are computationally or processing centered in their design. They consist of a central processing unit (CPU) which contains high-speed circuits (usually functioning at 10,000 to 500,000 operations per second) such as data and control word registers, timing generators, operational control equipment, and intracommunication switching mechanisms. Closely connected with the central processing unit, which adds, subtracts, multiples, and divides quantities (including the addresses of required control and processing information), is a high-speed memory requiring on the order of 1 to 20 microseconds to store or retrieve a computer word. Since small toroids (cores) of ferromagnetic material threaded with wires are usually used for this high-speed memory, it is usually referred to as core memory. Newer technology uses overlaid strips of thin film on an insulating base (substrate). This high-speed memory is generally from 2,000 to 32,000 computer words in capacity.

A computer word generally varies from 6 to 8 bits in length, for computers that must deal with information densely imbedded with alphabetical and decimal digits, to 18- to 72-bit word sequences, for scientifically oriented computers. The computer word is normally, but not necessarily, an integral multiple of 6 bits. Computer words or bytes frequently use 1 bit in the word-bit sequence for a process called parity checking. This process can detect a 1-bit error in the byte or word by use of circuits that check whether the total number of ones (or zeros) is an odd (or even) quantity. Any deviation from a preset condition alerts the operator and causes a "read again," "write again," or "stop" operation depending in what process the error is discovered.

The high-speed memory, in the large capacity size, can hold about one million bits. Depending on the processing being accomplished, a portion of the memory capacity must be allocated to the processing instructions (stored program), tables of storage addresses of data that are to be processed, and vacant "dedicated space" in which new results can be inserted or "chaining" references made between noncontiguously located but related data. The largest of these memories, if fully allocated to information, could contain the equivalent of a 20,000-word book or perhaps 500 to 1,000 library cards. Because of their high access speed these memories are expensive, about 50 cents per bit of capacity. For this reason, most data processors are equipped with auxiliary memories which store information more cheaply and from which needed information is brought into or returned from the high-speed memory as needed (or perhaps more correctly, when expected to be of use). Since these peripheral or auxiliary memories generally operate at autonomous data speeds of recording and reading out, they require special communication channels which buffer or compensate for the differ-

735-898 O—64——6

ence in internal computer rates and the auxiliary memory data rates. Even in the cases where no data-rate matching is needed, a channel for communicating read and write commands to the auxiliary memory is needed. Sometimes a time-sharing switching arrangement is provided to allow the central processor to service or utilize data flow from several auxiliary memories on an apparently simultaneous basis.

Figure 5 shows, in summary form, aspects of the technology that is available for use in auxiliary memories. The approximate purchase cost per bit of storage for some of these techniques is plotted in figure 6. These data include the cost of necessary read-in and read-out equipment, but not the central computer, and assume no manual handling of the stored material (that is, change of tape reels, disks, etc.) in order to achieve access to separately stored information.

Computer technologists often speak of read-only and read-write memories. Application of magnetic technology in most cases results in an ability to record information, to erase all or portions of it, and then to rewrite, at reading speeds, the same or altered forms of the original information. Photo-optical memory technology generally requires a re-recording of stored information in order to change it, although in some techniques under development a write-over capability may be achieved at speeds much slower than the read rate. Read-only should not be automatically considered a derogatory memory characterization since in many large files the percentage of change of recorded information is small over long periods of time (e.g. large library catalogs). In such cases, combinations of read-only, large-capacity memories with smaller read-write memories as addenda files are a possible solution to achievement of the high-capacity, rapid access features offered, for example, by the photo-optical technology.

Fundamental Aspects of File Organization

We have seen how the conventional data processor is organized. It has a computing unit (CPU) and a high-speed memory to service its data manipulation registers. Auxiliary memory units supporting the CPU and its high-speed memory are connected to them by communication switching devices and possibly data-rate buffering memories. Before considering the applicability of such technology to the automated storage and access of library bibliographic information, it is expedient to consider the fundamental constraints that operate in the design of optimized files.

As long as man has stored material, certain guiding principles have been inherent in the organization of his system of storage. First and foremost, those things, or analogously the items of information in the case we are considering, that are used (or predicted to be used) most often will be placed in the most convenient place of facility for use.

Designation	Recording material	Form of recorded information	Status
Magnetic: Tape, drum, disk.	Magnetizable coating on surface of plastic tape, metal cylinder, or disk.	Magnetization of regions of recording material.	Commercially available.
Magnetic cores	Ferromagnetic material made in toroidal (ring) form.	Direction of magnetization in toroid.	Commercially available.
Magnetic thin films	Overlayed "ribbons" of electrically conducting and magnetizable metallic strips.	Magnetization of regions of the thin film strips.	Recently commercially available.
Photo-optical	Silver halide photographic emulsions.	Opaque and transparent regions of film.	Operational, advanced development.
Magnetic-optical	Thin magnetizable optically reflecting film.	Polarization effect on reflected light from magnetized film regions.	Development.
Thermoplastic	Plastic tape—softened thermally during write operation by electron beam which distorts surface's optical properties.	Distortion of surface of tape, causing "lens" effects.	Development.

FIGURE 5.—*Computer memory technologies.*

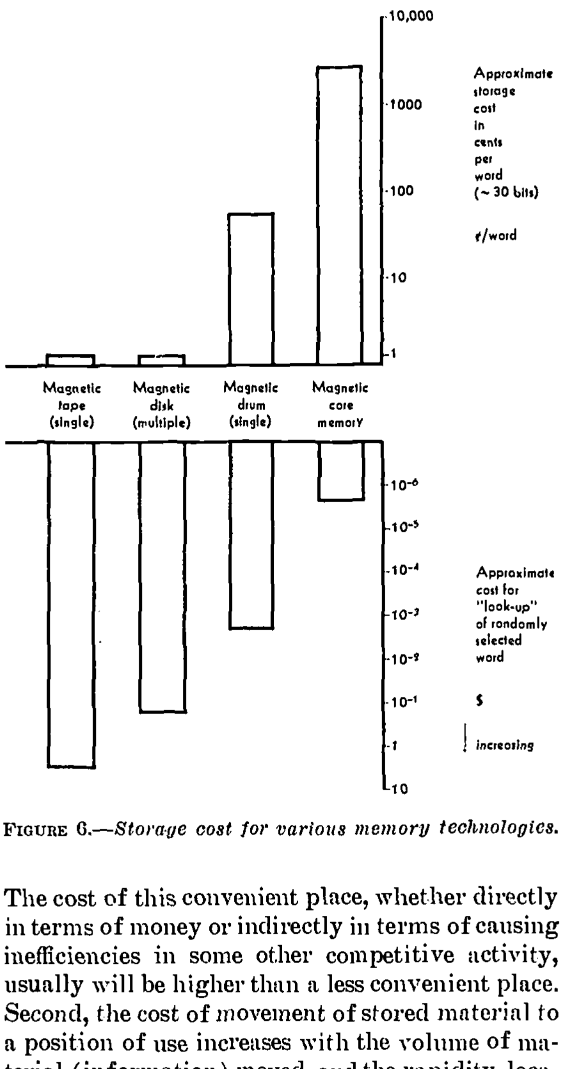


FIGURE 6.—*Storage cost for various memory technologies.*

The cost of this convenient place, whether directly in terms of money or indirectly in terms of causing inefficiencies in some other competitive activity, usually will be higher than a less convenient place. Second, the cost of movement of stored material to a position of use increases with the volume of material (information) moved, and the rapidity, location dispersal, and distance of movement required.

These constraints are not novel; librarians continually live with them. They become particularly acute, however, when the volume of information stored begins to cause undue expense (time) for patron or staff for access to descriptive items (cards) or holdings (books, etc.). Similarly, the definition of a convenient place for retrieval and use may change. This is the decentralization or branch library problem.

The designer of an electronic information processing system must face these same problems. Magnetic core storage has proven to be the most convenient storage place for data to be imminently processed or frequently used in the computer central processing unit. The larger memories of this type are of the order of one million bits in capacity and they can generally furnish to the processing unit the equivalent of something a little less than one average English language word each microsecond (one-millionth of a second). More dramatically stated, they can furnish on the order of one million English words per second for examination, alteration, comparison, etc., to the processing unit, although the latter may be able to do all these things only at an average rate of one-half to one-tenth this rate. The reason for this is that in each operation, for example, a comparison, two things must be moved from memory to special registers to compare the items and certain operating instructions also have to be retrieved from the working memory. Reference to figure 6 reveals as one would expect, that this "most convenient" storage is the most expensive for the quiescent holding of information even though it is least costly per lookup operation. Thus electronic systems designers are faced with the same problem as librarians—how to match a set of facility features of graded cost with the pattern of use of stored information.

Life would indeed be simple if one could categorize information to be stored in a library according to imprint date, time intervals, classification category, etc., and be reasonably assured that a sharp difference in frequency of the use of items in each category would occur. Unfortunately, with few exceptions, such is not the case. The distribution of use of information contained in segments of language, whether the segments are individual words or aggregates of words such as journal articles and books, is characterized generally by infrequent use of many items whose total use, however, is far from negligible. Now if convenience of access of stored items is not of concern (either to a data processor or a library patron), then requests can be accumulated and sorted (in computer parlance, batched and ordered) and access to the material can be efficiently accomplished for the benefit of the operating retrieval system, but unfortunately, not for the benefit of the user of the retrieval system, whether it be a machine or a human being.

One of the most typical use-distributions of language segments, familiar to linguists and workers on mechanical translation, is that of Zipf's law. (See item 15.) This states that if words of a language are listed in order of their frequency of use (or occurrence in large amounts of text) and are given numbers (called rank, r) starting with one for the most frequently used word ("the" in the English language) and increasing in assigned number as the use of each word becomes less frequent, then the probability of the use (p) or of the occurrence of each word is related to the rank of the word by the approximate relation:

$$p \approx \frac{0.1}{r}$$

Figure 7 shows this relationship and figure 8 illustrates its integral (the accumulative probability) and clearly demonstrates that, although a few words (of lowest rank) may account for 80 percent of the word uses, many other words must be available, although each infrequently, to account for the totality of English expression.

Similarly, librarians are familiar with the phenomenon described by Bradford's law of scattering which deals with larger segments of language, such as journals. (See item 14.) Here again, as figure 9 illustrates, the situation is similar—a large number of journals, each infrequently used, account for significant use.

One may inquire: Is this remaining fraction of the total use of words, journals, etc., really significant? Would it not be possible to cut off the storage of words, journals, books, etc., at some value of probability of use, e.g. for a given time period? The intuitive reaction of librarians against such a proposal has possible foundations other than experience. For example, information theory shows that the unexpected or least probable events carry the most information per event, even though information theory does not consider the value or utility of a quantity of information.

Investigations have been conducted on patterns of use of library material in order to determine more economical matching of storage facilities with frequency of use (regional repositories). (See item 3.) Here again, with the possible exception of information in the cumulative sciences, no sharp changes in frequency of use vs. imprint date, etc., occur.

We can now see another problem of large library automation emerging. How does one match automated memory technology with such distribution-of-use curves for file contents? Prior to further consideration of this matter it is appropriate to review automated memories and their search principles.

Automated Memories and Their Search Principles

There are basically two types of memories (not memory systems) categorized by the method used in placing information in the memory and retrieving it therefrom. The first of these are called absolutely addressable or extrinsically-addressed memories. Memories in this category usually require the specification of a numerical quantity, but not necessarily, for each "searchable" dimension of the memory. The term "extrinsically-addressed" refers to the fact that the address to particular locations of the memory is not based on any intrinsic or "contained property" of the information stored in the memory. The second type of memory is the content-addressable memory, also referred to as the integrally-searched memory, intrinsically-addressed memory, and the associative memory.

Prime examples of absolutely-addressable memories are the high-speed, magnetic-core memories used with the central processing unit on most computers. Most conventional magnetic tape, magnetic disk file, and magnetic drum storage units which are used as auxiliary computer storage devices fall in this category.

In each case, for this type of memory an instruction or command containing a "store" or "read-out" order and a numerical address must be given by the controlling device. In the case of the magnetic-core memories the instruction will contain a number which the memory circuits interpret as 2 memory coordinate dimensions which locate a stored computer word of fixed length anywhere from 6 to 80 bits long which is stored parallel to the third memory coordinate. The magnetic tape requires the specification of a particular record and file location along its length, and for the magnetic disks and drums specification of concentric or circumferentially located information tracks and angular sectors (segments of each track) is necessary.

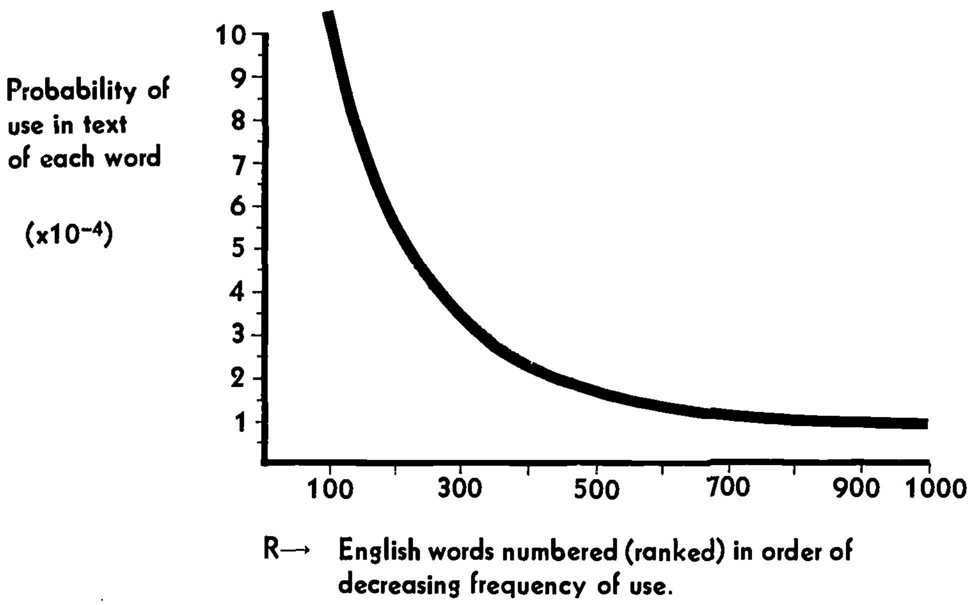


FIGURE 7. *Typical Zipf's law word-use distribution.*

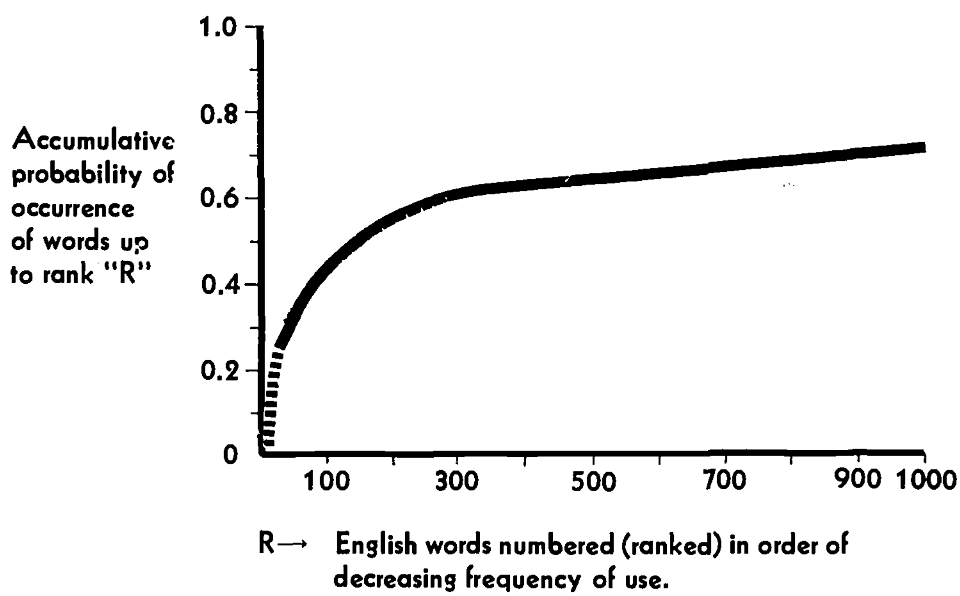


FIGURE 8. *Typical Zipf's word-use distribution: accumulative probability.*

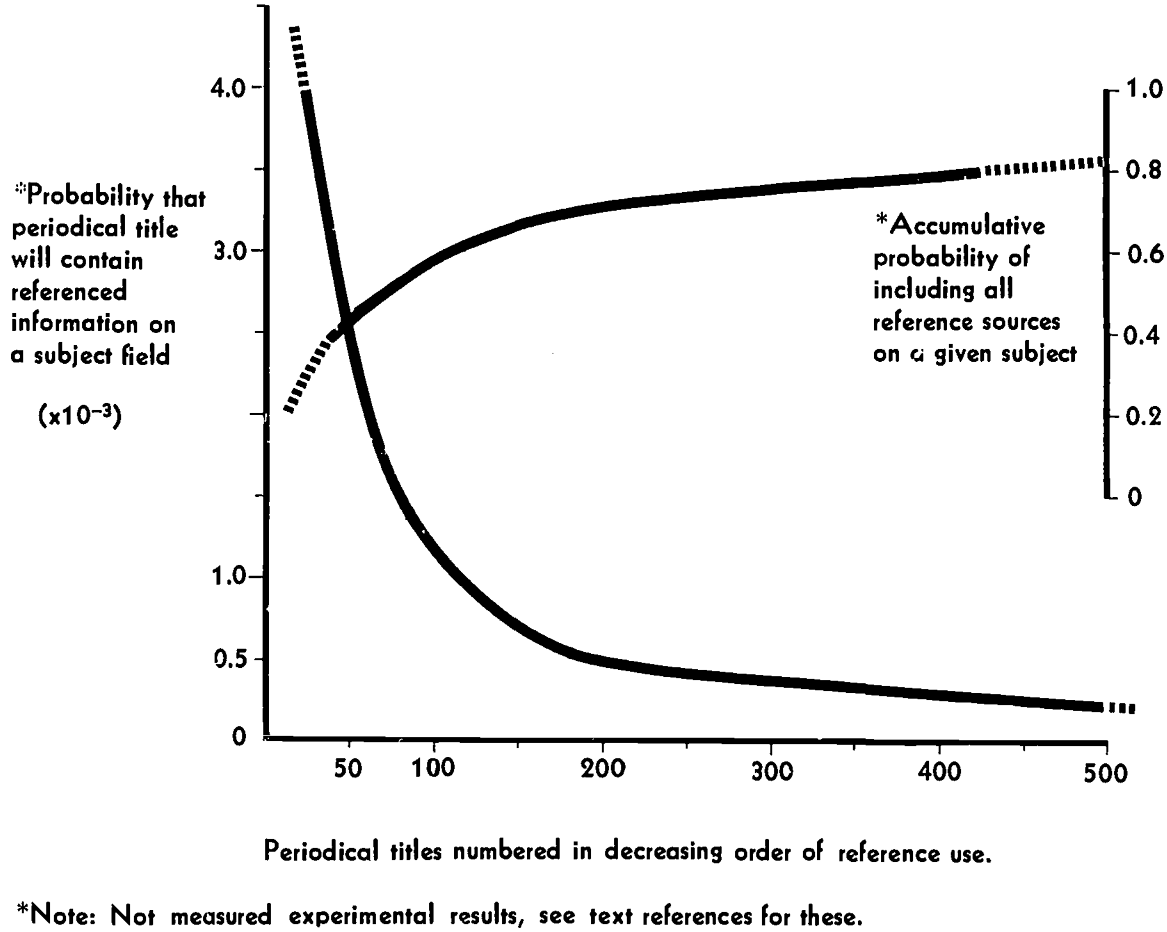


FIGURE 9. *Typical distributions of Bradford's law of scattering.*

The Content-Addressable Memory and Associative Memories

Information which is primarily expressed in language form is most naturally described and addressed in that form. This fact favors the use, for library information handling, of memories which are intrinsically or content addressable. Addressing of a memory by specification of some particular segment of the memory's stored contents would be of little advantage unless information related or associated with this specified content segment is also found. Accordingly, content-addressable memories imply some degree of associativity in their addressing structure and operation. This associative memory function and its implementation is discussed below.

Few content-addressable memories of an automated type are in existence. (See items 9 and 10.) One that has been successfully used in the operational translation of Russian to English and in automatic translation of stenotype code to English, operates by specification of a word, a number, a term, an indefinite sequence of words, etc. After such specification, the memory proceeds with a search of its contents much like one would look up the definition of a word in a dictionary. A word on an opened page (disk track) would be sampled; if this word comes before or after the word desired (say in alphabetical ordering), the pages are turned in the appropriate direction (tracks are stepped across) and a word sampled on each. This process continues until a page is found that contains a word that, in alphabetical ordering, is just beyond (greater than) the specified word. The memory then scans each word on the page sequentially until the specified input is found and its definition, translation, document or page number, etc., is read out. The search strategy and the ordering of "page" contents are such that the longest sequence of characters (and spaces) in the dictionary that match the input sequence of characters is found. This provides an indefinite-length content-addressing capability.

Definition of Associative Memory.—Associative memories in recent years have been proposed as a useful mode of storage and retrieval of information that has been recorded in digital machine-readable form. The term associative memory has many varied interpretations, almost as numerous as the proposed applications and the physical embodiments thereof. The term has been applied to adaptive-learning machines, parallel-search memories, vaguely defined information retrieval development goals, and computer-memory-address mapping devices, to mention just a few examples. To establish a common terminology as a basis of discussion, the term "associative memory" should be clearly separated from any implication that a particular hardware implementation or physical form of system is involved. Rather, associative memory is a label describing a criterion of memory performance. It defines a memory system with a capability of producing, or recalling from its stored contents, segments of information that are related to a specified item of information. The relatedness of output segments to specified items may be established by an internal processing by the memory system or by a priori input preprocessing, or both. Such a definition removes the erroneous implication that a particular device should be identified with the associative-memory function and admits rather that a number of possible configurations of hardware, and indeed hardware and people, could be used to accomplish an associative-memory function. The implementation in hardware form of the associative-memory function runs a gamut of possibilities ranging from research models of speculative cost and system utility to adaptations of standard-production computer components, whose utility (or lack of it) may be readily assessed.

Basic Methods of Implementing Associative Memories.—The implementation of the associative-memory function can vary considerably. At one extreme information to be stored is merely added to a file (e.g. magnetic tape, drum, or disk) after stipulating boundaries for its significant segments. Segment boundaries may be those of documents, chapters, pages, paragraphs, and even sentences, records, or fields. The associative-memory function is accomplished in this case by serially passing the totality of stored information past a retrieval statement and logically comparing each word or word stem of the file contents with the retrieval statement (or statements). Given extremely high-speed logic (compared to the transfer rate of the file data) and segment-sized buffering to perform the comparisons, a single pass of the total file would be sufficient to determine (and hence produce as output) those information segments in the file that totally or partially satisfy the retrieval statement components and conditions, including allowable masking and permutations of the latter. This mode of operation may be termed the "serial-search associative memory."

At the other extreme in functioning, the totality of stored information consisting of bounded predetermined segments (for example, sentences) is mapped into physically identifiable sections of memory hardware. These segments of information can be subcategorized in some way (e.g., as subject words or predicate words), to correspond to a similar categorization of any retrieval statement, the latter categorization dictated by hardware constraints. The specification of a retrieval statement in a suitable input-output register suffices to bring about the output of any stored segment that corresponds partially, if tagging or masking is accomplished, or wholly with the retrieval statement. This form of implementation is called the "parallel-search associative memory." As yet only hardware capable of handling computer-word segments and binary digit subcategories has been realized.

A method of implementation of the associative memory that falls in mode of operation between the completely serial and the parallel search approaches is the "integral-search associative memory" approach. In this method the information to be stored is processed once, prior to storage, in such a manner that any fundamental portion of it (word, term, phrase, date, quantity) that could possibly be of retrieval utility or significance, becomes a tag which allows initiation of a chained retrieval process to occur for any segment of information to which it is related.

Special Cases. A special case of the parallel-search associative-memory operation should be mentioned more for the sake of completeness than for its imminence of extensive practical use. This could be aptly called "weighed-network associative memory." In such memories, which are being proposed and studied in many research areas, the information stored exists as various states of excitation, or predisposition toward excitation, of networks of active and passive elements. In this class would be included many proposed memories being investigated under the following names: conditional probability machines, Cybertron, Perceptron, artificial neural networks, and automata.

Combinations. Of course, combinations of the methods cited above are possible. For example, serial search and parallel search could be combined for example, by using the parallel-search memory to hold the retrieval statement(s) and serially passing the total file by this for comparison. The flexibility of logic in matching may not be as great, however, as the use of programmed serial search methods.

The State of the Art in Associative-Memory Devices.—*The Serial-Search Associative Memory.* This mode of attempting to implement the associative-memory function is perhaps the most widespread in proposed or current application. Its usual manifestation consists of the use of magnetic tape or the newer magnetic disk files for storing the basic information (or in some cases a compacted form, abstracts, etc.). Standard computer main frames (core memories and computing registers) are used to retain the retrieval statement(s) and to perform comparison operations with the file data as it streams from the storage unit. The file data, of course, may be altered and restored or provided as output during the retrieval process. The retrieval operations are limited only by the programming effort exerted and the costs of machine time. Retrieval statements can be batched and usually are for economy. Unfortunately, if large files are to be examined for each retrieval operation, total file examination can rapidly become an uneconomical operation. In addition there appears to be no opportunity for human intervention in systems such as these to permit guidance during a particular search operation. Developmental improvement on this method of achieving the associative-memory function would appear to fall in the category of making the files bigger and the file transfer rates faster and possibly replacing the standard processors with special search-logic hardware which, though faster, would require limited-length mask and search registers with attendant difficulties and limitations in input microformatting (fixed word lengths, linking-bit tags, etc.) of the stored data. Generally speaking, this developmental approach seems to fall in the category of trying to defer recognition of the failure of standard data processing techniques in providing associative-memory functions. This mode of operation using, for example, magnetic tapes would, however, provide a capability for small-scale simulation of associative-memory functioning.

The Parallel-Search Associative Memory. Perhaps no hardware component development in recent years has excited operational information processors more than that of parallel-search memories. This has happened to the extent that both the memory developers and others have identified this mode of memory operation as being "the" associative-memory technique. Close examination reveals, however, that feasibility (ignoring cost) has been demonstrated at most for capacities of the order of 10,000 computer words. The largest parallel-search memory reported uses thin-film technology; other laboratory models have employed both magnetic cores and cryotrons.

The Integral-Search Associative Memory. A large class of information processing problems dealing with language relates to the problem of storage and retrieval of information segments whose probability of use is governed by a Zipf-type distribution. In this type of use distribution, common to most information retrieval activities, relatively few stored items account for about one-half the retrieval actions, but the remaining one-half of the retrieval actions are accounted for by a tremendous number of items which are retrieved infrequently. In this situation a memory capable of holding tens and hundreds of millions of bits of information, yet having access times in tens of milliseconds, has been shown to be useful and economical.

Memory Access

Before discussing memory or file systems further, some mention should be made about the detailed mechanisms of memory access. The access process, whether integrally addressed or absolutely addressed, can proceed in three basic ways. First, selection of a particular coordinate location (along a dimension of the memory) can proceed by a step-by-step passage, serial in time, along the memory coordinate by the read-in or read-out mechanism. This process in turn can be discontinuous or continuous; that is, the search can be serial or random. Second, the selection of a coordinate location can be made nearly instantaneous by use of a treelike circuit structure of binary switches which provide an input or output path directly to a memory-coordinate location. Third, the coordinate selection may involve the complete connection to all unit increments of one coordinate of the memory, or by connecting many circuit paths and read-record heads (by the aforementioned "circuit tree" mode of operation) it may read out of the memory many bits simultaneously along a given dimension. Any given memory may utilize all or only one of these access techniques.

A few examples should suffice to illustrate this point. Magnetic-tape auxiliary computer memories may record across the width of a ½-inch wide tape, 7 bits simultaneously, by use of 7 magnetic recording heads. This is enough to record 1 out of more than 60 different symbols along with a mark which is used to insure no error in the recording or subsequent read-out. The magnetic tape is given a linear motion in its long dimension (say 2,400 feet) and symbols (7-bit patterns) are sequentially placed along the tape. The linear density of these 7-bit patterns (characters, bytes) along the tape length is usually between 100 and 800 symbols or characters per inch. The density of bits across the tape is usually on the order of 10 to 20 bits per inch. Tape movement during recording or read-out will be on the order of several tens of feet per second. Magnetic tape is a good example of how read-out or search of one memory or file coordinate can be both continuous and discrete. When serial and continuous search of the tape reaches the end, or after a previous search is completed and a record that has been passed by is desired, then a rewinding of the tape must occur. Although this is usually done at much greater tape speeds than the recording or reading process, it still can add appreciable time to the average random access time to information on the tape. This access time is generally measured in minutes.

Magnetic drums and magnetic disk files, which are essentially competive memory embodiments, overcome the rewind problem by cyclically passing circular tracks of information under the read-record magnetic heads. Here the average random access time for comparable large-capacity files (10^8 or 10^9 bits) stored on drum and disk would run about 100 to 200 milliseconds. For smaller (several million bit storage capacity) drum files, average random access times can be on the order of 10 milliseconds. In the case of magnetic drums, information is recorded on a cylindrical surface by fixed electronically selectable or mechanically positionable read-record heads as separate closed circular tracks on the surface of the cylindrical drum. In magnetic disks the tracks are located as concentric circles on one or more disks and read-record magnetic heads are mechanically positioned along the radius of the disk for addressing the stored information.

In the future the technology for large, low-cost-per-bit, digital storage devices will probably use mechanically positionable selection mechanisms to a greater extent than an electronically selectable multiplicity of memory-scanning heads. There will be, no doubt, an exception from current practice in that they will use self-tracking rather than absolutely positioned scanning devices, since the latter appear to have reached a limit to their performance in large high-density memories. For access to random information on the cyclically scanned memories (for information within a track) the serial access time is one-half the rotation period divided by the number of times the same information is replicated (sectioned) in a distributed manner around the track. If track selection is accomplished by an essentially constant rate, two directional movement of the scan head between tracks, the contribution to the total search time is one-third the time it takes the head to move across the full range of tracks. This presumes no replication of information between tracks and a random use of the file contents. Electronic selection of a multiplicity of scanning heads (say one per each track) of course reduces the effect of track selection time on the total access time to a matter of microseconds.

Information Transfer Rates

No discussion of file storage and ac[illegible] would be complete without discussion of data transfer rates. Location of desired information in a store, e.g. a record or field of information, is only one-half the necessary complete cycle of retrieval. In a complete memory operation, a record or some segment of information is either transferred to the memory for recording there, or it is read out and transferred to some other processing or display mechanism. This transfer can occupy an appreciable fraction of the memory cycle time depending on the length of the information segment and also on whether the information is transferred simultaneously, many bits at once, (parallel read-out) or serially, one bit after another. Consider for example the case if read-out is accomplished serially at a rate of 600,000 bits per second. These may be accumulated in a 6-bit "byte" register until it is filled and then all 6-bit values transferred to another location (over 6 wires) at a rate of 100,000 bytes (or characters) per second. By this method, a device that operates slowly, for example, in recording a bit of information, but which has many parallel lines to its memory cells (recording positions), can receive information from a faster bit-rate device.

Transfer rates from magnetic tapes usually range (with special exceptions) from 60,000 to 600,000 bits per second. Magnetic disk files and magnetic drums operate at equivalent serial transfer rates of around 1-million bits per second. It is interesting to note that only when retrieval of long records is involved is the transfer rate directly significant in these latter devices. The typical library catalog card when located in a drum or disk storage memory would be transferred out at a one megabit per second rate (10^6 bits per second) in less than 2/1000 of a second.

File Access—the Man-Machine Interface

Data processing developments have yielded a plethora of computer input-output devices. These devices range from typewriterlike devices, for both input and output, to specialized-font page readers for input and high-speed multicase printers for output, both operating at hundreds of lines per minute of text. Recently there have been several developmental reports of voice input to machines and operational cases of "machine talks back to man." These two latter types have been based upon controlled or limited sets of voice, word, and message sets and imminent practical application to the handling of library bibliographic material should not be predicted. A library tracing (card) generator with typewriter input is shown in figure 10.

If a progressive approach is to be considered for handling the automated library man-machine interface it is probable that the "keyboard input and cathode-ray-tube display" console will be a favored approach. It is true that the electric typewriter in its various form (punched tape and magnetic tape operated, etc.) represents a possible file-user terminal as an interim technological measure, particularly for remote terminals where the economics of high-speed data transmission of graphics (images) to widely dispersed points may limit full

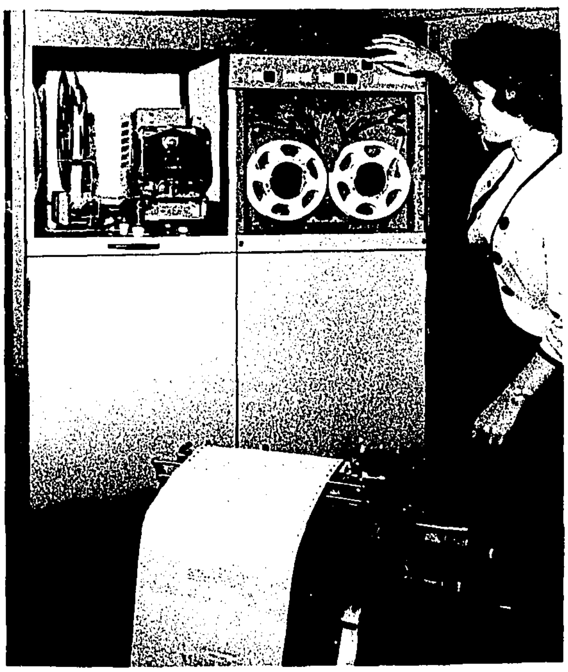


FIGURE 10. *Crossfiler—an automated library-card generating equipment.*

exploitation of the TV-like display capability that is inherent with cathode-ray-tube-type displays. In considering the possibility of cathode-ray-tube display and keyboard input consoles as a man-machine interface it must be noted that, up to now, these have been fairly expensive devices. It can be presumed, however, that if production quantities can be made on a few standard types, purchase costs for an adequate console for handling bibliographic information can approach $15,000 or less.

Two basic extremes in functioning are possible with such consoles. The first consists of providing each console with all the necessary character generation, display regenerating, logic, and buffer memory that is required for its operation. At the other extreme the console itself can be designed with a minimum of self-contained memory, logic, etc., and a central memory and logic unit capable of handling a large number of such consoles can be provided. Three consoles, representing different existing designs in this range of possibilities are depicted in figures 11, 12, and 13. In figure 11, the Electrada Corporation Model 408–2 console is shown. This console is a completely self-contained operating unit requiring only digital signal input and output connections (in the particular model shown, punched paper tape input/output media was specified and provided). The operator of this particular console can view a received message (up to 500 characters) on the upper half of the screen, perform any operation that he desires on the message with the keyboard by transferring all or selected portions of the received message to the lower half of the screen, and then can transmit the revised message. Alternatively, the operator can compose his own message, or he can recall from a console-contained memory any of 20 plus messages (which can be stored at will) and alter them, fill in blanks, etc., and transmit them. In this console, display characters (up to 63), edit symbols, and operational symbols are all generated within the console.

In figure 12, the Thompson Ramo Wooldridge TRW–80 Control Display Console is shown. This console also contains its own character, symbol, and display regeneration equipment but works in conjunction with a display-input buffer which provides an interface between a number of such consoles and any of several general purpose computers. This console is capable of displaying both line and symbol type information with a symbol size changing capability. Symbols can also be modulated in intensity, e.g. a particular symbol could blink or flicker, to alert an operator. This console has a capability for positioning displayed information under computer control. The operator can select particular displayed data, or the location for keyboard insertion of data, by means of a manually controlled cursor and a light pencil. The display screen is capable of displaying 2,048 symbols (about 340 English words) on the 16 by 12 inch CRT display area. Twenty-five lighted indicators inform the operator of the status of data processing sequences or modes of operation and a computer communication keyboard of 30 keys provides for operator selection of operational modes, etc.

In figure 13 an example of a display console is shown, the Itek Digital/Graphic Processor, in which the generation of displayed symbols, characters, drawings, etc., is completely computer controlled (the operator can select these or generate his own). Display regeneration on this, and similar, consoles is handled by a central shared memory unit. This unit uses both keyboard and "light gun" for communication of graphics or text with an automated system. Selection, alteration, repositioning, and multiple scale changing, are possible through use of a light gun, process selection keys, and a typewriter keyboard.

The particular consoles shown admit many variations in engineering and performance specifications and indeed are not the only kinds of consoles made by these companies or by others. Console design is intimately connected with the design of the overall automated system; therefore, a discussion of pertinent console considerations will be the next topic. None of the consoles illustrated is specifically designed for high resolution display of televisionlike text images as an alternative mode of operation. However, no fundamental engineering limitations should exist to prevent achieving such a capability if remote viewing of microforms is to become part of the capabilities of the automated library of the future. Each unit displayed provides more than 1,000 lines per display field and image quality in normal office lighting environments is excellent.

Basic Console Considerations.[16]—The basic functional components of an alphanumeric console of system pertinence are as follows (all may not necessarily be required):

1. Display
2. Hard-copy reproduction
3. Display marker control
4. Internal message storage
5. Process-control communication keys
6. Display symbol generation
7. Internal logic
8. Alphanumeric communication keyboard
9. Input/output interfaces
10. Automatic message manipulation

From a systems viewpoint each of these basic component functions interacts with the others and hence its characteristics cannot be independently specified, but for purposes of discussion they will be treated individually. In considering the desirable characteristics of consoles, the answers to at least four operational questions are paramount. First, what are the characteristics (form, size, symbol set) of the message entity to be displayed? The phrase "message entity" refers to the most predominant and important segment of information that the console user will desire to visually examine as a contiguous portion of text. Second, what

[16] This section on console characteristics was adapted from work by the author on Contract AF 19 (628)-10, sponsored by the Rome Air Development Center of the Electronic Systems Division, Air Force Systems Command.

FIGURE 11. *Electrada Model 408-2 edit/display console (Courtesy of Electrada Corporation).*

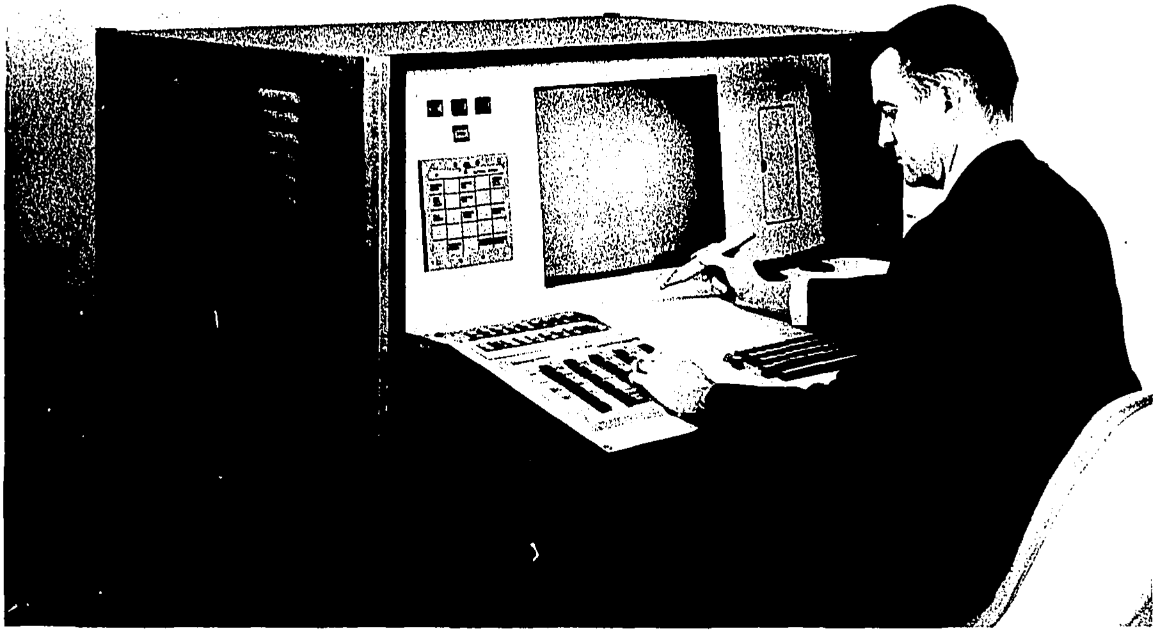


FIGURE 12. *The TRW–85 graphic control/display console (Courtesy of Thompson Ramo Wooldridge, Inc.)*

FIGURE 13. *Digital/graphic display processor.*

fraction of the total of the displayed message entities will be required in hard-copy form? Third, if display to an individual console user is presumed, is the rate of display of the "new text" to be matched to human reading rates or human scan and recognition rates? Finally, the fourth question, are trained keyboard operators to be the primary console operators?

Each of these questions is considered below in an attempt to derive conclusions having general applicational value with respect to the basic functional console components.

Console Displays. If a human operator is to continually compare or cross-reference the displayed message—with, say, an existing hard-copy version of the message entity (such as in proofreading) or to some well-known format such as a catalog card—the visual or mental cross-referencing would be greatly facilitated by exact correspondence between the format of the external or the "learned" message and the display. Similarly, if the console operator is viewing the display for the primary purpose of examining its geometric arrangement (such as in composing publication formats), then again the display format of messages should be allowed to correspond to some specific geometric format (i.e. characters allowed per line, tabular indentations, etc.). If this is the case, the input machine-readable message must contain function codes (carriage return, tab, etc.) that the internal logic of the console can interpret as display format instructions.

In many console applications the message displayed will be processed by the console operator primarily by examining the message with respect to its self-contained context. In these cases the value of formatting the display becomes doubtful and its occasional use as a geometric format-observing device can be handled by insertion, on the display, of interpretable symbols.

Display resolution should exceed 1,000 lines per normal viewing field, and brightness and range of viewing angle should be comparable with daylight TV displays so that operation can occur in acceptable interior illumination environments.

Can one specify a generally applicable message entity size and shape that would have wide application? If one specifies that the console display is to handle messages that consist of textual material the answer to the size question may be that obtainable on an heuristic basis; that is, the natural evolution of the paragraph (estimated as on the order of 100 words) as a message entity. With respect to shape, Belevitch suggests that normally occurring paragraph format has a theoretical explanation. (See item 1.) He proceeds on the assumption that the visual exploration of a paragraph of text would be most efficient if the number of words in each line (equal to a) and the number of lines in a paragraph (equal to b) is a minimum (that is $a+b=$minimum) for any specified paragraph size measured in words ($ab=constant$). This condition is satisfied for $a=b$; that is, the average number of words in a line is equal to the average number of lines in a paragraph. Belevitch indicates that this relationship appears to hold approximately, even for such paragraph-like text segments as public announcements. Presumably for library work the catalog card word capacity, with some allowance for growth in average size (about 300 words total), would be adequate.

Hard-Copy Reproduction. If the console application is primarily or solely one in which an experienced console operator is initiating a message entity and the generation rate is primarily operator limited (or is not of great concern), and a hard-copy reproduction for record purposes is desired, then for cost reasons (at least with present technology) the functions of display can be accomplished by a typewriterlike console. When typing errors or operator inexperience (and intolerance) is of concern, or the console function is primarily to assist human beings in reviewing or performing minor editing on machine generated message entities, then electric typewriters may be slow (most electrically operated typewriters do not exceed a rate of 20 characters per second). In this case, electronically produced displays should be favored. With the cathode-ray-tube (CRT) type of display at least two options exist for obtaining hard copy from the console. One of these involves conversion of the electrical coded message into hard copy by facsimile, electric typewriter, or printer methods. An attractive alternative from a cost viewpoint would be use of an optical-photographic method (not necessarily silver halide) working directly from the CRT display image.

If appreciable hard-copy output is to be obtained from a console primarily for record purposes, it must be remembered that the cost of photolike reproduction materials is proportional to their area; minimum linear dimensions of copies, consistent with legibility, is advisable.

Display Marker Control. On the question of combining, in a console, the functions of operator display and hard-copy reproductions, consideration should be given to "multiple choice" operation. In most console applications it will be advisable to keep the user's input operations (such as typing words) to a minimum. In such a mode of operation the capability of moving a pointer or marker to various portions of a displayed message and then pressing a "process key" to indicate specialized treatment of the message segment indicated is a valuable feature. However, it must not be presumed that a typewriter with hard-copy display could not also have such a feature although it could involve slower operations than an electronic display would provide, such as a reverse paper feed and typing head movement.

The question of message entity display rates is a difficult one to answer on a general application basis. Three operational modes, at least, should be distinguished in considering this question. First, there is the type of operation where batched sequences of messages are being paraded in front of the console user for the purpose of his reading or editing them in toto. In this case, the read-in period should not be accomplished at a slower rate than 10 to 15 characters per second which is the normal reading rate, and preferably it should be much higher to avoid waiting periods at the console that are annoyingly much longer than message entity reading periods. Second, there is the case where some group of message segments are read into the internal storage of the console and the console user must scan these rapidly, e.g. for a search and recognition process and, possibly, a subsequent editing and tagging action. In this case, if maximum convenience to and efficacy of the console user is desired, then the display of each message entity must be as easy and rapid as turning the pages of a book. Third, there is the case where a number of stored pre-prepared messages are contained in the console and called onto the display so that the operator can in some manner alter or add to them prior to their transmission. The throughput processing time for such messages and their frequency of use would dictate display rates in this case. However, the use of "canned" messages usually presupposes an objective other than just error minimization in their repeated generation, and here again display of such messages should require a small fraction of the time that would normally be taken by their being newly composed each time they are needed.

Internal Message Store. No general rule can be given concerning the magnitude of the message entity storage that should be incorporated within consoles. This would be a parameter highly dependent upon the total system functioning involved. It should be noted, however, that if such messages are to be used as a readily available repertory for an individual console user, such as in a cataloging operation or in an intercommunication system, then the user would begin to require a table of message storage references if the number of messages were much more than 30 to 40 (assuming that no easily remembered, ordered or hierarchical, relationship exists between the stored messages). The question of console message storage of a repertory of message entities is essentially one of system costs. If many consoles (employed by users concerned with nonidentical but possibly overlapping stores of message entities) are to be employed in a console-computer system, then the use of centralized memory should be considered up to the point where the costs due to required central memory access times and communication and switching (including computer interrupts) makes it advantageous to place sufficient memory at the console terminal to handle the high-use message entities.

When a console is being used by a human being to scan a block of message entities and the user is selecting and compiling a group of these, then the previously mentioned ability of obtaining hard copy of displayed items could greatly reduce the required magnitude of console electronic "scratch pad" memory. Despite this, it would appear important, to avoid undue manual reinputting of portions of the compiled information (say for further system processing), to be able to store on an electronic "scratch pad" basis, at least up to one display-size of information.

Process Control. Without question, if the console operators are experienced, a large number of special keys to control both console operations and associated data processor operations would present little difficulty. Viewing a display screen while typing input data to a console is not a difficult operation for an experienced typist or machine operator. To the relatively untrained operator, sequences of viewing a display and then transferring eyes and hands to a nonstandard process-control keyboard could be disconcerting. In the case where consoles are integrally connected to information processors, serious consideration should be given to use of a limited set of process-control keys and maximum use of multiple choice displays for selecting and sending choices of words and their operational symbols which the data processor can interpret, within allowable sequences of processes, for the particular operation being performed.

Designers should give serious consideration to merging all keys that effect the display into the same keyboard. For example, display marker controls may be incorporated into a segmented space bar or the backspace key of a normal alphanumeric keyboard configuration. A foot-switch and indicator light could then be used to indicate whether marker control or message editing generation is the function being performed.

Display-Symbol Generation. The symbol set required for console display must generally provide for the following:

1. Reproduction of the textual material to be processed.
2. Indication of special operations that have been, or will be, accomplished on the displayed text by output printers or attached data processors.
3. Indication of the type and location-effect of certain console operations.
4. Alerting the console operator.
5. Notification of the status of console processing.

The trade-off between the use of symbols that serve purposes 1 and 2 above is considerable and can result in lowered console costs. For example, if source text is generated in electrical code form and contains code symbols for uppercase shift and lowercase shift, the console may be required to provide only two display symbols to indicate to the console user whether a character from the set of single-case alphabet characters in the console display repertory represents a capital letter or a lowercase letter. The same trade-off can be used to reduce the internal logic necessary to recreate the exact source text format by providing symbols that indicate paragraph indentations, etc. As was mentioned previously, if format editing or proofreading is a prime console use, then this trade-off would not be desirable.

An important decision in library automation concerns the number of alphabets that must be used on console displays. System cost considerations argue strongly for standardization on the Roman alphabet even though the internal file storage contains notations of other alphabets for special publication uses.

In the display of symbols that indicate where on the display some console operation is to occur, or has occurred, usually a simple "sweep" intensity time-gating action that provides an underlining along with a suitably labeled process key is sufficient, and character-generation codes and equipment are not needed. In the case where a message entity is held on the display and various segments of its are stored or transferred to an editing region, some consideration should be given to the insertion of editorial notations such as "permanent" dots under the characters that have been processed. This would be of particular aid, for example, in cataloging operations where message entities may have repeated phrases and words which make it difficult for the console user to locate his prior actions in the message being processed.

Internal Logic. The relationship between the incoming data stream and what is displayed has an appreciable effect on the complexity, and hence the cost, of consoles. The internal logic may be so constructed that not only is each code byte or character code examined for suitable display-symbol generation or console action (e.g. stop receiving) but sequences of character codes may even have to be examined. This latter case is particularly true if the source text is encoded in a set of bytes that are insufficient to convey by each byte a unique representation. This is the familiar case of using a 6-level code (maximum unique representation, 64 items) to represent a symbol set consisting of all alphabetical characters upper and lowercase, punctuation, decimal numbers, and special characters. (In this case about 80 symbols are represented by the 6-level code.) The question of how many information levels of code consoles should be designed to handle is not one that can be generally answered. Seven levels (information, exclusive of the parity bit) have been increasingly employed in equipment to generate and code text. The resulting 120 plus symbols that can be directly encoded would prove useful if computer programming symbols and other communicating specialties are to be handled without special source-character coding and decoding.

Alphanumeric Communication Keyboard. There has been some experience with nonstandard typewriter keyboards on consoles, and the reaction has been essentially negative. Although there is no such thing as a completely standard keyboard with respect either to the set of characters, special symbols, and punctuation, or their position, console keyboard specifications should require as close a correspondence as possible to the keyboard of commercial electric typewriters. Consideration should be given to combining display marker keys with the alphanumeric keyboard.

Input/Output Interfaces. In the case of on-line consoles connected directly to a file or file processor, the interface problems and considerations are at least as numerous and varied as the types of processors that may be involved. A problem that is immediately faced is the initiation of a computer "interrupt" or "call for console service." If one console is involved then presumably the existing interrupt schemes of most commercial data processors can be used. If a number of consoles are involved, possibly simultaneously, identification of the interrupt may become a problem. Consoles generally, although not necessarily, can be considered as asynchronous devices, and some input/output data-rate buffering would be required as well as a consideration of matching "word" structure. Presumably, the processor would be capable of examining the console output data stream for instructions inserted by the console or the console user. The reverse situation, where the data processor is "instructing" the console in automated operations, should be kept to a minimum from a cost point of view and instructions should be provided to the operator via the display.

Automatic Message Manipulation. There are numerous functions that can be accomplished on an automatic basis within a console or externally to it. For example, if the console user deletes words in a message entity it should be possible to close up the created gap on the display, or conversely, to fill the gap with a "delete" code and symbol which later accomplishes the same thing on an output printing device. Insertion, deletion, and transfer of message segments in increments down to the level of a single character or symbol are necessary if editing for spelling is envisioned as an operational requirement. The decision as to whether the display should hold a message entity while it simultaneously allows a message editing or composing task on another display area is primarily an operational one. If the console user is to use the displayed message as a basis of composing a different message entity (e.g. an index card, information retrieval summary, answers to a problem) then dual input/output message display is probably desirable. If the primary process is one of making minor editing changes, then single display may be suitable, if, however, the input or unmodified message can be sequentially displayed should the need arise.

Human Engineering Aspects.—Certain human engineering aspects of consoles, such as process key placement and keyboard characteristics have already been mentioned. If a console is to serve as an effective transducer between man and machine, its design should be tailored to the convenience and capabilities of the expected user. Display legibility should approach that of good quality printing. Where possible, the mode of operation should be one in which the operator has a minimum of operations to perform and these should be visually and manually constrained to limited physical areas. Small conveniences such as desk space, text-holding devices (near the display), and end-of-line signals all contribute to the user's impression of the console as a usable and desirable tool. Safety features for protection of both the console and its user should be mandatory for any design.

On-Line Console Uses.—Three major categories of on-line console uses can be identified, although combinations and variations are possible:

1. Text editing—format and content
2. Message composition—copy and insertion
3. Tutorial interaction and file access

735-898 O—64——7

In editing text that has been converted to machine-readable form, computers (and lexical processors) can perform many editing functions automatically. The entry or non-entry of information in certain designated portions of the message entities can be detected. Common spelling errors can be corrected, and ambiguous spelling situations can be flagged. For those cases where some ambiguity remains in the machine-edited text, a console operation involving human assistance is in order. Such editing can involve both rearrangement of the geometric layout of the text as it will finally be desired in the output as well as alteration of characters, symbols, punctuation, and words of the text.

Message composition can be achieved in two principal ways. In the first of these, an operator views an input message, extracts or paraphrases portions of it, and essentially creates a new message. An example of this would be the use of the console by a document cataloger or abstracter. In the second, the operator inserts keywords, names, or addresses in "canned" or pre-stored messages. The two modes are not clearly separable; for example, a message cataloger or indexer could extract sequences from an input message entity but insert them into a prepared message format that consisted of a few headings such as accession number, date, source, or keywords. The insertion mode of message composition could involve inserting fixed data, data locations, etc., into a computer program as well as straightforward secretarial operations.

Tutorial interaction and file access cover the rather broad spectrum of console uses that pertain to man-machine communication. Console uses in an automated information storage and retrieval system, where the "machine" attempts to assist the man in organizing his query, and automatic teaching machine uses would fall in this category.

File Storage and Access System Considerations

The most important aspect of automated system design is the kind of service that is to be provided by the system. It is on this point and the accompanying cost considerations that opinions about how to design an automated bibliographic information system will differ. To the author's knowledge, nobody has proved that finding a reference to a monograph or serial containing a desired segment of information in x minutes instead of $10x$ minutes is generally worth y expended dollars. Similarly, it is doubtful if it can ever be shown on a firm quantitative basis that instead of finding n references to a specified topic that finding an $(n+1)$th is generally worth y dollars. It is the author's moot conviction that automation of the handling of bibliographic descriptive information will parallel the usefulness and the success of automated improvement of telephone communication. Retrieval of information contained in a library is essentially the establishment of communication between authors and library patrons.

There is no reason to believe that existence of recorded knowledge prior to a particular need for it should restrict our ability to gain access to it at a later date. Thus, an author may write a book which is filed away in a library before there has been an expressed need for the information it contains. Once a user needs that information, however, he should be able to gain access to it just as quickly as if he were to consult personally with the author by telephone. If this premise is correct, then the goal of automation should be to help the library patron find useful communication linkages and to do this with the maximum speed and convenience that technology allows. In a sense there is an implicit assertion in the foregoing statements that timely availability of information is by far the most important factor in the value of information.

Granted that there is a telephone communication parallelism, does this aid in determining how the automated bibliographic system should operate? It does, for conversation generally is a bilateral process of giving and receiving information, not just a posing of questions and a responding with strictly appropriate answers. The author is well aware of the vast literature on information retrieval, library classification, etc., which describes methods upon which any inanimate bibliographic system must base its response to queries.

Consideration of the system applicability of file storage and access technology need not be based, however, on a priori selection of any one of these bibliographic methods. One can start by assuming that any initial library query will be based on the use of language, and further that the initial expression for desired information should generally consist of not more than 5 or 10 significant terms (names, words, etc.). These terms may have polarized relationships (i.e. acted on, co-incident with, etc.) although their value in the light of cost of implementing is questionable. This point as well as many others is discussed in two review papers, items 6 and 8.

The automated bibliographic system must respond to this initial query by some indication of the acceptability of these query terms, e.g. not found, too large a response expected, etc., and with a suggested list of related terms. These suggested related terms can be based, for example, on coincidence in titles, co-occurrence in authority catalog cards, or comembership in a segment of hierarchical subject structure. Based on average works per author, average number of tracings, etc., an estimated response of significant terms on the order of 25 to 100 suggested terms could result from an initial set of 5 to 10 terms.

Higher quantity responses could be controlled by tutorial-like suggestions. A request concerning Shakespeare could result in a suggestion to specify authority cards, drama, sonnets, etc. The library user could select additional terms from the bibliographic response and reiterate this process until he has satisfied himself that he has an adequate set of query terms. Reinsertion of the final selected terms would elicit a response as to the number of catalog entries each term would yield. The user would then be able to try combinations of joint occurrence of terms on a trial basis and would obtain a display indicating the number of catalog entries (monographs and other library holdings) that would be identified. This operation would complete the term-search process. Entry of the final term selections along with their specified logical relationships (specified joint occurrences) would provide at the console a display of the appropriate catalog cards for the user's review or hard-copy printout.

In terms of file storage and access technology, what would the service outlined above require? First, the term-memory would require a memory technology that was primarily content-addressable. The addresses found in this memory by specification of languagelike terms (including names) would yield "see also" numerical addresses (catalog card numbers) which would contain the cross-referenced terms. It must be possible to add to this term-memory both completely new term entries and insertions of new item numbers to established term entries since new catalog cards will contain previously posted names and terms. The second memory, the catalog-card-memory, would contain the equivalent of library catalog cards addressable by number with significant retrieval terms and names annotated within them. This memory would require additive properties on a sequential or accession sequence basis only. Some logical processing to provide for the selection of terms from the full catalog or for subsequent application of retrieval criteria such as imprint dates, book size, etc., would be required.

Two important questions remain, concerning the capacities and the throughput rate capabilities that would be required of these memories. With respect to capacity it is estimated that several hundred thousand English word stems and phrases would be required in the term-memory along with names of authors and certain spelling variations of both terms and names. For a 5-million-item library with 1 million authors—assuming that any given retrieval-significant term occurs in the bibliographic material, e.g. catalog card text, at a frequency of one in every one million words—a capacity of between 10^9 and 10^{10} bits would be required for each of the memories.

Throughput refers to the rate at which a file storage and access system can handle queries as inputs and provide responses. Consider each query as a 2-step process, first as the insertion of terms with their subsequent lookup which yields catalog card numbers from the term-memory. If 5 terms were inserted and if each lookup read-out is accomplished in the average time of 10 milliseconds then 20 such query steps would be handled in a 1-second period.

If the 5 initial term lookups each yielded 10 catalog card numbers there would be a total of 50 catalog numbers to be looked up in the catalog-memory. If these in turn took 10 milliseconds each, then one-half a second would be required and only 2 such queries could be processed in 1 second. If response delays (peak) of several seconds can be tolerated then more users could be serviced within a given time period.

The desired speed of response to queries (query throughput rate) affects the complexity of system design. Methods of congestion theory in telephone systems have been exhaustively treated (see item 13), but it is doubtful whether in most library systems the "large population" results of telephone theory can be applied. Recent work indicates that the selection of methods for automated system servicing of query queues is not a particularly critical one. (See item 2.) The relationship of internal activity in a system procedure, such as outlined above, to console user activity would require some thorough investigation before realistic estimates could be made. It will probably be necessary to "over design" automated library systems of the man-machine interplay type, until more experience is gained with such systems.

Technology vis-à-vis Automated Bibliographic Information Handling

The development of technology for processing information that has a predictable use and a numerical character and the work on new technology for handling information having predominantly language characteristics have resulted in system file storage components that can meet the requirements of automating bibliographic information handling in large libraries. The achievement of such a goal will require design of memory-centered systems capable of handling natural languages, rather than processing-centered systems which place severe constraints on input data preparation. Considerable attention will have to be devoted to the development of man-machine interface equipment and to system throughput considerations.

Bibliography

1. Belevitch, Vitold. Langage des machines et langage humain. Bruxelles, Office de publicité, 1956. 119 p.
2. Eisen, M. On switching problems requiring queuing theory in computer based systems. IRE transactions on communications systems, v. 10, Sept. 1962: 299-303.
3. Fussler, Herman H., *and* Julian L. Simon. Patterns in the use of books in large research libraries. [Chicago] University of Chicago Library, 1961. 1. v. (various pagings)
4. Goldman, Stanford. Information theory. New York, Prentice-Hall, 1953, 385 p. (Prentice-Hall electrical engineering series)
5. Hartley, R. Transmission of information. Bell System technical journal, v. 7, July 1928: 535-563.
6. Herner, Saul. Methods of organizing information for storage and searching. American documentation, v. 13, Jan. 1962: 3-14.
7. Illinois. University. *Coordinated Science Laboratory.* Human performance in information transmission. Urbana, Ill., 1955. 69 p.
   Reprinted Feb. 1959.
   PB159-800.
8. Jaster, J. J., B. R. Murray, *and* M. Taube. The state of the art of coordinate indexing. Washington, Documentation, Inc., 1962. 256 p.
   Contract NSF-C-147.
   ASTIA document no. AD 275 393.
9. King, G. W. Photographic techniques for information storage. *In* Institute of Radio Engineers. Proceedings of the IRE, v. 41, Oct. 1953: 421-428.
10. King, G. W. Table look-up procedures in language processing. pt. 1. The raw text. IBM journal of research and development, v. 5, Apr. 1961: 86-92.
11. Pierce, J. R., *and* J. E. Karlin. Information rate of a human channel. *In* Institute of Radio Engineers. Proceedings of the IRE, v. 45, Mar. 1957: 360.
   See also:
   Karlin, J. E. Reading rates and the information rate of the human channel. Bell System technical journal, v. 36, Mar. 1957: 497-51C.
12. Shannon, C. E. A mathematical theory of communication. Bell System technical journal, v. 27, July, Oct. 1948: 379-423, 623-656.
13. Syski, Ryszard. Introduction to congestion theory in telephone systems. London, Oliver & Boyd, 1960. 742 p.
14. Vickery, B. C. Bradford's law of scattering. Journal of documentation, v. 4, Dec. 1948: 198-203.
   See also:
   Stevens, Rolland E. Characteristics of subject literatures. [Chicago, Publication Committee of the Association of College and Reference Libraries, 1953] 10-21 p. (ACRL monographs, no. 6)
15. Zipf, George Kingsley. Human behavior and the principle of least human effort; an introduction to human ecology. Cambridge, Mass., Addison-Wesley Press, 1949. 573 p.
   See also:
   Mandelbrot, Benoit B. An informational theory of the structure of language based upon the theory of the statistical matching of messages and coding. *In* Symposium on Information Theory, *2d, London, 1952.* Communication theory; papers read at a symposium on Applications of Communication Theory. Edited by Willis Jackson. London, Butterworths, 1953. p. 486-502.

CONFERENCE SESSION III

Mechanization of File Storage and Access

RICHARD L. LIBBY
Itek Corp.

Problems of Information Handling

Throughout my paper I have attempted to expose those who are not familiar with some of the computer technologist's parlance to these terms and to give a rather cursory but, hopefully, comprehensive coverage of the mechanisms of file storage and access. Since this is a tremendously complex subject, I could only touch on the major features of mechanization that might be of interest to librarians. I begin my paper by pointing out that mechanized information handling has had its greatest success with the handling of information that is expressible in quantitative amounts and with information whose use can be predicted with some certainty and which can be categorized with considerable expectancy that any subsequent use will fall within the assigned categories. Librarians deal with information of slightly different characteristics. In many cases they cannot predict who will ask what questions. This type of information handling problem is shared by people attempting to mechanize the handling of management information and intelligence information. One cannot in an a priori sense easily or successfully categorize the information to be stored in these mechanical graveyards. I am not certain that we will find all the solutions in the future, but we must seek methods of mechanizing language-type information that will allow a greater freedom of access to the potential user. Hopefully this can be done without posing a tremendous intellectual burden at the input end.

Information that librarians deal with is such that one cannot pick out segments of it and say, "This is used so much percentage of the time, and I will allocate its storage to this type of equipment; this is never used or used very infrequently, so I will put it in a regional depository," and so forth. As a matter of fact, the frequency of the use of words and the occurrence of useful references on a given subject in journals both behave according to either a hyperbolic or a reciprocal distribution where relatively few items (in the case of words, perhaps 100 to 200 words) correspond to 50 percent of the total use. The remaining 50 percent of use is distributed over a tremendous number of words or journals, any one of which may be used very infrequently, but the total aggregate use of all these other things can add up to significant use of stored information.

The Content-Addressable Memory

Now, this has a relationship, which I confess my paper does not point out very clearly, to the fact that the normal mode of operation of numerical data processing equipment employs a method of absolute addressing; that is, one maintains tables that lead to blocks of information in the store which are called for as the need arises. What is needed in the handling of bibliographic material is a technology that employs memories that are content-addressable; that is, one does not specify an absolute, coordinant record, but one specifies segments of the information and this in itself leads to the address of related and pertinent information. This may not be at all surprising to people who are familiar with card catalogs. This is exactly what you do if you look for an author's name; you go to the place in the file that is ordered by the author's name. In present machine technology, particularly as an outgrowth of the old numerical processing techniques, you would have to examine the first two characters of the name and determine the number of the tape on which that name could be found. Then if there were a tape switching unit, you could call in the block of information from the appropriate tape and then examine this in detail. My paper outlines the status of the so-called content-addressable memory (also called intrinsically-addressed memory or associative memory).

Consoles for Library Access

There is a problem of communication with data processors. In the past, use of numerical data processors was essentially by means of a batching operation because these machines cost from $60 to $600 an hour whether they are used or not. There is a tendency, of course, to make certain that they are used every possible second in order to get a lower cost per operation. For the computed payrolls and that sort of thing one can schedule the use of the computers very well. I personally question whether this is the proper approach to foist upon librarians or anybody else who is running an individual service organization. I think that if technology cannot provide accessed information when and where the user would like it, then we must go back and do our homework before we technologists come to librarians and say, "This is what you ought to do."

Consoles can vary from simple typewriterlike devices to units costing on the order of $100,000 to $150,000. Most of the military-sponsored developments have been in cathode-ray-tube display consoles coupled with an electric typewriter, with some storage within them to regenerate the displays and so forth. These consoles have incorporated within them many functions that are useful for drawing pictures or inserting text and editing it by means of the keyboard, and, as a result, these consoles have cost about $100,000 apiece. Such costs are, of course, unthinkable for libraries. As a matter of fact, in addition to the cost of the console there are tremendous programming costs to allow the use of 10, 20, or 100 of these on-line with the computer; there is the time sharing problem and so forth. I venture the thought that there is no reason why consoles adequate for automated library catalog access cannot, with a production run of 100 to 200, approach something on the order of $15,000. I base this very moot point on the fact that the console of the type I have in mind would essentially consist of an electric typewriter keyboard, which might cost in production about $3,000—possibly a great deal less—and which would use a character-generation scheme of about 120 characters, with perhaps 7 allocated operational codes, and have a monoscope or a digital generation of characters. I feel that this latter portion of the unit should cost on the order of $5,000. The necessary logic that would go with both the keyboard and the display generation should not be greater than $5,000. Let me point out, though, that, to my knowledge, nobody has developed and put such a console on the market. I cannot explain this except for the fact that there has been no coordinated and standardized need expressed for a simple console; therefore consoles have all been custom-designed, and, hence, expensive. The display and console characteristics are treated in some detail in my paper, providing, I hope, a checklist so that anyone who is interested in this problem can use my paper as a takeoff point.

File Arrangement

I, too, propose mechanization of the card catalog by a dichotomy, one part of which is a dictionary file consisting of many subdictionaries: author dictionaries, subject dictionaries, and so forth. I propose this arrangement of the catalog primarily to keep the throughput problem within reasonable limits. The other half of the mechanized catalog is the complete bibliographic entry, i.e. the main entry, in machine-readable form and essentially in accession number order. I propose the use of as many terms as possible as entries in the dictionary files. For those terms (such as Napoleon or Shakespeare) that become almost useless as specific retrieval items, special tables should be displayed to the user to help him narrow his search. Indicia such as the color of a book or the dimensions of a book may be useful for the retrieval of books previously used. This would be helpful when the only thing one remembers is that "I had a yellow book that dealt with the frequency of words in the English language." When the final set of catalog cards in machine-readable form is retrieved after a search of the store, then logical processes can be used to discriminate items in this final set of cards by criteria which are not very selective for early stages of the search but which would become effective after one has narrowed the search.

The Feasibility of Library Mechanization

As far as the feasibility of mechanizing library operations is concerned, I think that today we are at the point where many of our standard commercial products can mechanize that portion of large libraries that deal with housekeeping, e.g. keeping track of the in-process, the not-on-shelf, the circulation, and similar activities. I have no question that it is possible and that it could be done in line with present costs, with the possible benefit of having up-to-date information about the current location of bibliographic items, how many there are, how they are used, and so forth. With respect to automation of the descriptive media called card catalogs, authority catalogs, and so forth, I feel that the technologists have demonstrated in the laboratory the technical feasibility. However, I feel that some very direct and enthusiastically sponsored effort would be needed to put this into what one might call "off-the-shelf" capability.

The Librarian and Information Control

MORTIMER TAUBE

Documentation, Inc.

The 3 by 5 Syndrome

I would like to add to what was said about the card catalog, because, after all, the card catalog is the chief mode of access in a library and that is what this paper is about: access to bibliographical information. I would certainly admit that the library profession has done well with the 3 by 5 card in handling monographic publications. But we must not forget that the library gave up the problem of handling the journal literature to the scientific societies. The societies did not handle this literature with 3 by 5 cards. After the war, when there was a great volume of report literature, the Atomic Energy Commission made an initial attempt to handle it with 3 by 5 cards. This attempt has probably been one of the greatest bibliographical failures of all times. Sending out millions and millions of cards which remain unfiled in libraries around the country generated a new bibliographical term: shoeboxes. The question was: How many shoeboxes of unfiled cards do you have in your library? The 3 by 5 card, although it has had great use within the monographic field, has certainly not solved the bibliographical problem.

You may say that as librarians you are not concerned with report literature, that your concern is only with organizing a library in terms of monographic material. If you do say that—and I hope you will not—the chances are that the modern librarian will pass out of the picture as the paleographer did some time ago. I teach a course at the graduate library school at Columbia University. Recently I took my class to IBM to see the machines in operation and to see various programs being run. We saw one of IBM's motion pictures which showed plant x where a man had certain files he was cataloging, using ordinary cards. It also showed a picture of the man's door. It said "Librarian" on the door. The man went inside and there were all these cards and all these files that he worked on. Gradually this man got more and more intelligent and brighter and brighter, and he mechanized this part of his work and that part of his work and another part, and he went from punched cards to small computers to large computers. Then IBM showed another picture of this man's office; the word "Librarian" was gone, and on the man's door was the title "Director of Information Services."

I remember that a number of years ago this problem was first presented to the Special Libraries Association. SLA said, "This is no problem for industry; this is only a government problem; this is only a Washington problem that you fellows are trying to fool with." Soon it became not a Washington problem but a special library problem and an industry problem. What has happened, in many cases, is that the librarian who refused to be concerned has not become the Director of Information but has been placed under a Director of Information. So the problem as I see it is this: in the future either the librarian will make his peace with the modern world—that is, he will take this new technology and make it a detail in his operation—or, if he does not, what the librarian does will become a detail in someone else's operation. We may feel that this is not going to affect the academic library and that it is not going to affect the true research library, but that was my point about the paleographer. How many of you academic librarians employ paleographers? There was a time when you could not be a librarian unless you were a philologist and a paleographer. These men went away. Now if you librarians think solely in terms of the management of serial records, vertical files, and 3 by 5 card catalogs, you will go the way of the paleographer.

Comments on Measures of Information

Now to come to Libby's paper. The paper deals first with a measure of information, a mathematical measure of the number of bits of information that would be in a message, in a catalog card, in a page of text, and so on. This is an important question for the machine man because every bit of information that he stores costs some fraction of a dollar to put away and to get out again. If he can determine mathematically that he can record a certain amount of data in a more compressed form, a more compressed number of bits or digits, he is saving money. I did feel, as I read this, that the librarian could be excused if he felt that the logarithm of the number of bits and so on was a little bit beyond his immediate concern. He is willing, I think, to accept these details about the amount of information, or how things are coded, or how one gets the maximum information in the smallest area, from the information theorist. What the librarian wants to know is what this means in terms of the number of questions that he can answer in a unit time. He wants to know not necessarily how many bits or how many digits are involved in a particular reference question but how many reference questions he can answer in a given time. One of the ways in which the computer person can give this information to the librarian is to calculate the amount of bits that have to be processed. Here we must find an interface which I think we do not have now. In too much of the literature the computer man writes about speeds, number of bits, amount of information, and so on, rather than about the number of inquiries that can be handled per unit of time. The latter, which the computer man could calculate if he would go one step further, would make his material of more immediate concern to the librarian.

File Access and Structure

The paper also discusses computer access, and how the store is used to answer questions. Here, I think, I must relate this to the paper by Patrick and Black, which discussed in its appendix two arrangements of the store: inverted and linear. Now this is obviously concerned with the access to the store, and I would like to consider both of these problems together. Access is determined by the structure of the store, and structure involves two different things: physical structure and logical structure. The physical structure relates to the type of file—that is, tape, disk, and so on. If the data are on a tape and you want something that is in the middle of the tape, it is necessary to start at one end and unroll it. In other words, this particular physical structure has a constraint which requires a certain type of linear access. If the file is on a disk, you can take a reading head and move it to a certain line on the disk and find the data directly. On a drum storage you can run reading heads across the top and rotate the drum and find the data that way. In a core storage, where wires run to each core, you can address the core directly without any physical movement. These are the types of physical structures which determine how you get access to the material stored in the machine.

You have all heard about so-called random structures. Well, random is a bad word and we ought to get rid of it because what it ultimately means is ordered. It is a piece of technical jargon. Random means an equal time to get at any particular part of the store. To develop this further, the equality results when a system has reached maximum entropy and has no structure, so that the chance of one thing happening as compared to anything else happening is equal. Actually, a random store is a store where one can definitely specify an item and can go right to it; it is the opposite of linear.

It turns out that these ordered stores have a rising curve of expense: the paper is cheapest, the disk is a little more expensive, the drum is still more expensive, and the core is still more expensive. The more immediate the access, the more direct the access to an address, the more expensive. Therefore, one has to design a system that takes into account these physical differences in terms of the frequency with which the information is needed.

In addition to this problem of physical store, there is the problem of the logical structure of the system, and this involves such concepts as the principle of ordering. In other words, the logical structure of an author file is that it is ordered alphabetically by the author's name; a subject file has a logical structure in which there are subfiles arranged under each subject. Now in the computer art there has been considerable debate as to whether or not you have to structure a store in the same way as you structure a card file. That is, should you prefile the material under different subjects so that when you look for something you do not have to examine the total file but can go directly to what you want in the computer just as you do in an ordinary card file? As computers have gotten faster and faster and as we have been able to store material more and more densely, it has been suggested that we do not have to structure the file logically, because we can just store data as received, and the computer has enough power to answer questions by scanning the total file. Of course, when this is said, it is usually also mentioned that this procedure only becomes economical when one batches. In other words, since you must go through the total file to find any item, you decide to go through the total file to ask as many questions as possible during the trip.

According to Libby's paper this is perhaps the wrong way to go about it, because the library patron will never stand for his inquiries being batched. The patron will require from the new mechanized library the same direct service he gets from the present card catalog, and therefore we must supply him with a console or some means of interrogating the system directly without his waiting to be batched. Now the idea that any large library is going to be arranged efficiently in a system of direct access so that questions will not have to be batched, so that anybody can approach the store by console, seems to me difficult to accept. I realize that the art moves along very rapidly, so I am not going to say, as I have said about other things, that it is impossible. But I question the realism of telling the librarian: You are not going to have to change your way of doing business; you are not going to have to batch; you are not going to have to do so-and-so, because we're going to supply a console to you which is going to interrogate your file and give you immediate answers, regardless of the size of your file. My feeling on reading this part of the paper was that this was not realistic. This is my summary. The topic is now open for discussion.

General Discussion

SWANSON: The idea of a console does not necessarily imply that you cannot batch; it is just a question of the response time of the console. For example, if you had a 1-minute response time, you could batch all queries that come from 200 consoles during that minute. There may well be some intermediate kind of system here that relieves some of the direct-access strain on the hardware and permits batching over a reasonable response time.

WARHEIT: I agree. We must emphasize the difference between a minute and two minutes and the millisecond range. There are a lot of milliseconds within a couple of minutes.

KING: When you're talking about batching, you presuppose that you have a simple-minded question and that you will get only one answer. Well, Libby, I think, realized that the questions are not going to be simple—there will be stupid questions for which there are no simple answers, and to get any satisfactory answer out of the system, it, with the help of the querier, has to set up a search trail. So the answer is going to come after a long sequence of interrogations to the members of the system, maybe 100 or maybe even 1,000. Obviously you can't batch those.

HEILPRIN: I believe that what Taube said has some validity. I do not think one can talk about the response time of the console independently of the size of the collection. As the size increases, the response time almost certainly has to get longer. It may not be a linear function, but it will certainly get longer. The question of the size of the collection is very important and cannot be erased even with consoles and these split files.

BUCKLAND: What additional data, other than that on a library card, are to be stored in this way? I can't think of a very complicated question to ask of the data on a library card. While you people are thinking about converting files how about thinking about putting a little more data in the file? Then I can think of more complicated questions which ought to justify the use of a console. Right now it is a poor second to a book index made from the same data.

DUBESTER: I would like to emphasize the point that King made about the necessary dialogue making batching difficult. Reference librarians know that 9 times out of 10 the person who asks the question doesn't know the question that he should be asking; he is approximating the question that he wants to ask. If you have faith and just batch his question, the probability is that you are going to answer the wrong question. This suggests that there should be a dialogue between the system and the inquirer.

TAUBE: Dialogues with catalogs, in my experience, are not statistically very important. Now I will admit that there is sometimes a great ambiguity and uncertainty in the mind of a searcher concerning what heading he should use, but there are other ways of handling that. For example, we developed a system called "analog search." In other words, you can ask for a search of logical sums, if you wish, by specifying 10 terms and then saying, "I'll take the item if it has any 9, any 8, any 7, any 6, any 5, any 4, any 3, any 2, any 1 of the terms." We can handle that question in a batch because it's a formal question. Now if you say that we cannot succeed in library automation until we supply a question and answer between a store of *x* million books and every individual who comes into the library, I despair of the realism of this type of approach. But then I've despaired before and the world goes on.

WILLIAMS: It just occurred to me that one of the basic problems that seems to be facing the machine people here and the librarians is the relative magnitude of the files and the information that the machines have to handle, but when there is a much smaller store you can get quicker access to it at relatively cheaper prices in terms of frequency of use. One of the things that we now know about libraries, but so far in this discussion have forgotten, is the following: most of the material in libraries is very infrequently used. This would suggest that if librarians could define more accurately the materials which are frequently used and the types of questions most frequently asked, access to this material could be automated and you could have fast access to it; the rest could stay on the 3 by 5 cards, and delay in access to it could be accepted because this would happen infrequently. This kind of an approach might be a practical one for utilizing the present limitations in capacities of the machine.

TAUBE: If we could predict our problems it would be a lot easier. The difficulty is we cannot. My teacher, Alfred North Whitehead, used to say that you could burn half the books in the British Museum and nobody would know from now to the end of time that you had burned them. The only problem is, which half?

LIBBY: If we technologists are saying to the librarians that we can allow you to provide essentially the same convenience of service that you now achieve through these bits of pasteboard by mechanizing, then I think the technologists should just leave. If one talks about automating the bibliographic control of the library, then it must be for the purpose of providing this dialogue kind of operation. I cannot conceive of letting a user submit a question and then, after batching, the next day or an hour later answer the question, only to find out that he would have been able immediately to modify the search trail if a librarian had asked him a question or even said, "Do you mean this or that?"

TAUBE: One practical answer to that is what we call in our organization a man-machine information system. We would never think of putting any question, whether simple or complex, directly into the machine. The question passes through an interpreter. In many cases, where the questioner understands what you have, there is no requirement for dialogue—there might be just some substitution of terms. In the cases where the user does not know anything about the system, he is asked about his question before it goes into the machine and the dialogue takes place outside the machine.

PATRICK: I would hate to see our thinking constrained by assuming that we had to have communication built into the consoles. There are a multitude of problems that can be solved now, with presently available equipment, to start our learning and to build up this large file. We need the file built up either way. The communication mode, this conversation mode on these consoles, is a highly experimental laboratory device right now. There may be one or two military operations going, but these are very, very few; they are expensive; they are out of sight on price. This is a "gimmick" today. There are several experiments going on—Carnegie, MIT, Ramo Wooldridge, Rand Corporation, and SDC—where people are playing with these devices. These devices are not what I consider commercially available, reliable, 24-hour-a-day, useful working tools that we know how to handle. This is something I can think of, but being able to deliver it reliably is not anywhere near in the same class as doing payroll applications on a computer.

TAUBE: The danger of this approach is that by emphasizing the console and the dialogue, you take away from the librarian his obligation to formalize his problem and that is what he must do. If the librarian waits for the console and assumes that when it comes he will engage in dialogues and not have to formalize his problem, this will delay mechanization.

Let me give you one concrete example. We run two systems for the Cancer Chemotherapy National Service Center of NIH. One system deals with the action of ordinary drugs on tumors in mice; the biologist has been able to formalize this. He has said, "If the results are so-and-so, let me see them; if the results are not so-and-so, keep them in your machine; I don't want to see them; print them out when the experiment is over." He has given us formal instructions, and he sees maybe 1 out of every 1,000 test results that come into the machine. There aren't enough biologists to look at them in all of NIH.

On the other hand, as part of the same program we run a series of tests on endocrinology in which they require bio-assays before they do the tests. Now on this there is no agreement as to how the compound should act. The endocrinologist has not been able to formalize his problem and the machine doesn't help him, because he has to look at everything that comes in. Now you can say that what we ought to have for him is a dialogue, but it is much simpler to give him the whole business and let him make up his mind what to do with it. He realizes his situation is bad compared to his conferees and he is trying to state a set of rules, to formalize his problem. When he does, he will get the same benefit from the machine.

SWANSON: I don't agree with your fear that the use of the console is going to take away from the librarian the responsibility for formalizing his processes. I think that it, in fact, forces him to do just that, and without such consoles he is going to be mechanizing a system in which he is still keeping for himself the functions he is now performing in nonautomated systems. I will concede that any kind of mechanization will, of course, have to be preceded by formalization, so I don't particularly consider that a strong argument either for or against consoles.

I cannot understand Patrick's concern over the issue of reliability of consoles. To be sure, they are a good deal newer than computers are. However, it would be astonishing to me if engineers couldn't overcome reliability problems that now exist. If you are working on a time scale of a few months, I am sure this is of great concern, but people who are planning mechanization of large library systems are probably going to do it on a several-year time scale. The consoles that exist today have been generally custom built; to a degree they have been experimental. As soon as a genuine, marketable requirement can be established, I am confident that they are going to be mass produced with the same degree of reliability that computers have.

ORNE: I am one librarian perfectly willing to accept the fact that the machine people can design a machine that can handle all of the countless numbers of units that we need. This paper describes how information in card catalogs can be put into such machines. It has just been said that we would have to formalize our work if it were to go into machines. We *have* formalized our work. One thing is not in the paper, and this is the one thing that we are all looking for—a discussion about deeper analysis of the materials for library users. We can go to the library catalog and conduct this kind of question and answer business, or we could put it into a machine so that this question and answer business can be conducted either mechanically or visually; the only variation is in the amount of time it takes or the amount of expense one is willing to put into it. These papers offer some very ingenious solutions for ways and means of incorporating material into the machines. We know it can be done. The main question still remains: How does this take us further than where we are now? We probably would go with you, and eagerly, if it took us somewhere further. And whether at comparable expense or expense that is foreseeable to be paid in 10, 20, or even 50 years, we might start planning for it. But you have not shown where it will take us beyond the point where we are now. This is my sole argument with machines.

LIBBY: Your statement that the paper does not indicate how the automated catalog system could allow you to do much more than is done now with present methods is a valid criticism. However, I believe the paper does express the thought that the insertion of terms at a console and the use of a dictionary process to respond to each of these inserted terms to retrieve and display the pertinent catalog entries (which either have these terms as constituent elements or are related to these terms) would represent a greatly expanded capability over what you have now in the manual operation of the card catalog.

ORNE: We have the same thing now on cards; we have see references and see alsos.

LIBBY: No, you are presently more limited. As a librarian, Orne, I believe you will admit that the entry into a card catalog seeking a term does not guarantee the retrieval of all of the cards that may have included this term as a significant element. The catalog allows you to go to see alsos to the extent that on an a priori basis you have made an intellectual decision that there should be a see also card. I would venture the guess, however, that this is quite limited compared to what could be done, in principle, with an automated catalog system.

MINDER: May I suggest a rather elementary, but nevertheless real, example. The Census Bureau recently supplied us with the 1960 census on tape. It is possible, with the aid of the computer and the reference librarian, to get more data more quickly from the 1960 census than was previously possible. These census data are used over and over again by the students and faculty, and it is possible in our program to determine how frequently certain questions are asked. These questions will be put on a shorter program which will run more quickly. This is an example of the kind of thing we are talking about, although it is very elementary.

DUBESTER: If one concludes that an automated catalog will do about the same things we do now, then possibly the question "Why bother?" would be appropriate. However, there is this to observe also: the catalog, whether one likes it or not, is central to every bibliographical operation in a library. When the order clerk makes up a requisition, when the book comes in with an invoice to be cleared, when the preliminary cataloger makes a preliminary catalog card, when the cataloger prepares the final copy, and finally when the card is printed, the same information is generated over and over again. With automation, once information is entered in machine-readable form it is part of the system and you do not have to do the same thing over and over again.

A second advantage is being able to be self-conscious about the experience of the system, with respect to the kinds of questions asked, the number and type of subject headings that are applied, and the throughput rates of various parts of the system. One of the reasons that librarians have not been able to answer questions about costs is that they really haven't known how many times significant functions are performed in the library. They haven't been able to do the analysis which a machine system can do.

This leads me to another point. The thing that librarians must save in the future is not money as such, because money is, after all, an expression of a social value judgment at a given time and a given place. The real shortage that we are going to experience is the skilled manpower which is getting more and more scarce. We are not able to find the searchers, the typists, the people with language skills, and the catalogers. This is where we are going to be in a critical position. If we can do the formalization, optimum utilization of skilled manpower is one of the benefits of automation, with or without consoles.

TAUBE: Just for the record, we have had some exchange here and we have had two papers. Now as I understand the two papers, the mode of access recommended in the first paper is a linear file with batched inquiries; in the second paper it is console access. I had assumed that this meant random, now I'm not sure. Could this be clarified?

LIBBY: It is *not* linear.

PATRICK: State of the art means all things to all people, and this is the trouble here. State of the art to the men in the laboratory is 10 years distant to me. That is the difference in the two papers.

BUCKLAND: I want to comment on Orne's question of what could be done that is new with the console. I think that, rather than use the word console, we should use Swanson's definition which is "intellectual access to the information." Now I understood Orne to say that the cataloging process was a formal one. I would contend that, in the sense of Taube's definition of formal, that subject cataloging is not a formal process.

TAUBE: Nor is descriptive cataloging. Since 1938 librarians have been fighting about the rules: obviously cataloging is not a formal process.

BUCKLAND: It is not a formal process, and the reason I think it's not is that its basic medium of communication or interaction is language. Language by its very nature is an ambiguous form of expression. Now if you want to be unambiguous, you develop a branch of mathematics; you can say things precisely with a branch of mathematics. The thing that is wrong with it is that you can't say anything new. You develop a branch of mathematics which encompasses a certain scope, and to go beyond that you have to develop another one. Now in contrast to that you have language which, somehow, as ambiguous as it is, always allows you to express something new. People communicate with it all the time, and they don't do it on a one-shot basis. They say something, get a response, and they continue back and forth. Now this intellectual access to the file is going to be an analog of this type of communication; it may involve lots more than is now on catalog cards.

EDMUNDSON: I would like to defend Libby's paper; I think it was responsive to the initial request placed upon him. It does describe a state-of-the-art view. Some of the questions have caused me now to wonder if we are not facing the same problem we have often faced before. I would like to make it very clear that we should keep in mind the distinction between the computer and the program. It has been very fashionable to talk about a chess-playing machine, a chess-playing computer. What plays the game, badly or poorly, is not the machine, but the program. The librarian's programs will have to be designed for his needs. The consoles described in Libby's paper are accurate representations of presently existing hardware. They may not have the programs that meet your requirements at the present moment.

HEILPRIN: While listening to this discussion about consoles I tried to formulate a definition of a console. All search is identification, but in a library search we do not identify a *thing* until the end of that search. Until we get to this final act of identifying the thing we identify classes. Now to identify classes we use a kind of switching system. We find classes common to several larger classes by intersection; this is the basis of deeper indexing and of intersecting terms. However, to find the original terms, we first enlarge the search by taking a logical sum. This is what a thesaurus is. A thesaurus is a preliminary enlargement of classes to ensure finding more usable classes for intersection later, which cuts down on what Taube has called the man-machine dialogue.

Now what is a dialogue? A dialogue is only a series of intersections, logical sums, and negations suggested by the trail in the catalog. A see also is either a logical sum or an intersection. Thus, by console dialogue we really mean the following: How much of the logic of search can we predict, i.e. program, and how much do we have to extemporize? In general, the greater the association built into the system, the less the necessary dialogue and vice versa.

LIBBY: I feel that my paper has been identified as being in the category of the "blue-sky." I would refer the readers, or potential readers, to page 78 where I point out that a typewriterlike device would be an acceptable interim console. I thoroughly believe that if one is progressively planning an automated library one has to admit that the reproduction of remotely stored microimages by TV-like presentation should be taken into account as a future possibility for a library. Therefore, I tried to outline the kind of display that should eventually be used in an automated catalog system. However, I repeat that as an interim measure this console could be an electric typewriter in its various forms.

One other point: a dialogue does not have to be the posing of a structured question. It can be the insertion of simple terms, much as we use a book index. It is true that logical intersections and such can be accomplished as a result of the insertion of terms and the response from the file. But I am not proposing the use in the near future in a library of the "blue-sky" type of effort such as fact retrieval and so forth. I feel that a book-index kind of operation could be expanded to be very useful in a mechanized or automated catalog for bibliographic information.

TAUBE: I was asked to say something more about the notion of "formal." Here are a few simple examples from library situations of what I mean by formalization.

Every librarian who uses a dictionary catalog knows that when the number of cards behind a certain heading in the catalog gets too big you subdivide the heading. If you have too many headings under U.S.—HISTORY, you then subdivide it by U.S.—HISTORY—WAR OF 1812, or if that gets too big you use U.S.—HISTORY—WAR OF 1812—BIOGRAPHIES. There is a rule of thumb as to how these things are divided.

Now suppose you are going to do this with a computer, and the computer has been instructed that if there are so many cards under a heading it should subdivide them. The computer says, "*How* many?" Then you say, "I don't know, so many!" The computer says, "Give me a number!" That's what I mean by formal. It's as simple as that. We have worked out such a rule in one of the particular systems which we are struggling with. If we have a heading RADAR and another heading AIRBORNE RADAR, the computer counts the number of headings under AIRBORNE RADAR; if there are less than a specified number, it eliminates them and puts these items under RADAR. If there are already too many under RADAR, it counts the entries, takes them out, and puts them back under AIRBORNE RADAR. The point is that this is a rule whereas for the first 2 years of working at this thing we had people going through the file marking, writing, and saying: "Take these headings and put them here and take these postings and put them there." The moment we worked out a rule we put it into a computer, and the computer does it. And that's what I mean by formalizing the problem.

That is subject cataloging. Now to turn to descriptive cataloging with an example I have used many times; many of you know it. The three great national libraries of the world are in Paris, London, and Washington. For the one in Washington the official heading, according to the rules of the ALA and the Library of Congress, is *U.S. Library of Congress;* the one in Paris is *Paris. Bibliothèque nationale*, not *France. Bibliothèque nationale.* The one in London is neither *Great Britain. British Museum*, nor *London. British Museum*, but *British Museum.* Now here you have the three national libraries; the headings are all different.

CLAPP: How do you explain this?

TAUBE: You have a rule. You have an exception to that rule——.

CLAPP: That was a rhetorical question!

TAUBE: All right. The point is that this is the kind of thing you can't tell a computer to do. Some individuals have decided this on the basis of a rule, an exception to the rule, and an exception to the exception! If you want to get payoff from the computer you make a rule, and that's all I mean by a formal procedure.

CLAPP: Speaking of your numerical rule, I was once a member of the Decimal Classification Editorial Policy Committee and we were trying to decide on what basis to expand the Decimal Classification. After some rumination we came to a rule exactly as you described. The rule was that if there were more than 20 books in the Decimal Classification catalog under one item, subdivide. I bring this up only to mention that when this rule was published and got to England, the British Decimal Classification people objected; they said every book ought to have its specific number, even though there was only one item under that number. Now this is the opposition here, you see, which needs to be resolved.

TAUBE: Libby states that his paper ranges from the tutorial to the speculative, and he expects that the speculative part will occasion much disagreement. One of the major problems that confronts the entire scientific establishment of the United States is the question of how far ahead shall research be undertaken. I have in mind basic research, and it turns out that a number of people in Government have been looking at the kind of research supported by the Federal Government.

In some areas there is no serious problem—if you want to build so many miles of road or if you want to build an airplane, you either build the road or you don't, or the airplanes fly or they don't, or if you want to get a missile up in the sky, the missile goes up or it doesn't. Now in these types of technological work we do have standards, but when it comes to basic research, measuring its validity becomes a very difficult problem indeed, and determining how much money should be spent for this kind of thing becomes a very difficult problem.

Let us bring this down to the problem at hand: the argument for or against the console. I am concerned at this moment with what the librarian should do about this problem. As I watch my audience, I can see the librarians reveling in the fact that the machine people are arguing among themselves. This seems to remove the onus or the burden of decision from the librarian. They can sit back and watch the machine people argue as to which is the best way to go and they can say, "Until the machine people make up their minds, I don't have to do anything." Now if librarians follow that course, by the time the machine people make up their minds there will be nothing left for the librarian to do.

ANGELL: I make this remark with great reluctance because it is characterized by complete triviality, but if I understand you correctly, you have just said that while the machine people are arguing and making up their minds, all librarians can learn to be paleographers.

BRISTOL: I have a question regarding the text of Libby's paper, page 86. There is this statement: "One can start by assuming that any initial library query will be based on the use of language, and further that the initial expression for desired information should generally consist of not more than 5 to 10 significant terms (names, words, etc.)." This seems to me to be dubious in the light of ordinary reference service. We usually don't get 10 significant terms or anything like it.

LIBBY: The 5 or 10 reflects an uncertainty on my part. I would suspect the maximum would center about George Miller's magic number—7. Human beings tend to characterize, at maximum, any given concept of thought, off the cuff so to speak, with somewhere between 5 to 10 descriptive items. I can't back this up with any citations, but, at a console, I believe that you would not expect the initial entry of terms describing what a person is groping for to exceed 5 to 10.

WARHEIT: In a survey made in the AEC, over 90 percent of all the inquiries were three terms or less.

TAUBE: I think that most questions are much simpler than people suppose.

DUBESTER: We do not know whether people pose questions in three terms or less solely because the present system is designed to accept such questions, or whether they would provide seven terms or more or less if a different system were available. In other words, we really are not prepared to make this comparison.

TAUBE: It is recognized that the index to *Chemical Abstracts*, until recently, was the most scholarly, the most complete, the most detailed, of all scientific indexes. *Chemical Abstracts*' subdivisions and their modifications go on to about the fifth subdivision. By the time you get to the fifth subdivision, that is modification under a heading, you select one out of a million items. In other words, the amount retrieved by that last subdivision is usually one abstract. Now this indicates that asking for a product of seven or eight terms is probably going to screen out information and is probably too specific a question even for very large files. Does anyone want to comment?

LIBBY: I attempted to set an upper bound. I am pleased to hear that the initial entry of terms averages about three. This would certainly help the throughput situation in the mechanized or automated catalog.

TAUBE: This is an area where we would be better off if we had some specific numbers. As many of you perhaps remember, at the beginning of the study of machine methods one of the major problems discussed was superimposed coding—the question of how many codes could be put on an IBM card or on an edgenotched card. The reason why this work turned out not to have the impact it should have had on the library profession was because, in order to calculate the order of superimposition, one has to know how many terms there are going to be in the question. You can stand a certain degree of superimposition if somebody asked for a 7-term question, but if they ask for only a 2-term question with that degree of superimposition you'll drop the whole deck. Therefore, this whole discussion of superimposition has disappeared from the library literature, largely because we don't really know how many terms people use in asking for material.

WARHEIT: There is one dangerous element here, and this is the fact that when the inquirer approaches the librarian he is often apt to generalize his question for fear that he will not be understood. I have a feeling that in talking to the console, to the actual store, he might be much more specific. At Oak Ridge we were often asked to "send everything you've got on it." When we went back to the original requester we found he had a much more specific question.

PATRICK: It doesn't make any difference how many terms you give, you can't get too much, because we'll guard you against this. If you get over 50 items potential response, we'll say, "Look, you will get more than 50, are you sure you want them?" You'll come back and ask the question again.

TAUBE: I'm now going to use a dirty word that hasn't been used in the whole meeting—uniterm. When we first made a manual uniterm system, we had certain ideas concerning the distribution of postings. We knew as we counted the postings that accumulated under various headings that we had certain safety areas and certain troublesome areas. What we realized was that headings with less than five postings were too specific to be good retrieval points: the file tended to branch out into one posting under an almost indefinite number of headings. We realized that we also had headings which were not good retrieval terms because they retrieved half of the library. We knew that there were two things to do: for headings with few postings a check would find that these terms either were synonyms for terms that were already in the catalog and which therefore could be eliminated, or that they were too specific and you could post on a higher generic level and eliminate them. Of course, if it were a genuine new term it would remain unchanged. Similarly, terms with many postings were too general; they retrieved too much, and therefore they were examined to see whether or not they could be subdivided into more specific headings. Now I bring this up at this point because we knew all this theoretically, but we could never do it. For 6 or 7 years this whole idea was observed in the breach until we went on computers. Then this was duck soup, because we could count and we could organize and reorganize our file everyday so that it could be of maximum utility to the retrieval operation. This is what I mean about the computers; if you can formalize and put in a number, the computer can plan the distribution of postings in the store.

ANGELL: What does this have to do with the user, the public? What do you mean by something being too specific? There are at least two aspects to specificity: the specificity that is possible, which is a function of the system or of the language, and there is the specificity which is desirable. Now what is too specific?

TAUBE: This seems to me to be what Verner Clapp was pointing out when he said that the English refused to abide by the statistics. When you design a system like this you'll always have people say, "I'm not interested in statistics, this is *the* heading. I'm not interested in accommodating my indexing system to your machine, I am interested in the user and his point of view." I submit that, as librarians, the only thing we know about the user is that the more consistent our apparatus, the easier it is for the user. This it seems to me is the important thing about a computer system; if we can tell a computer how to set up an index, we can certainly tell the user how it has been set up. But if we can't tell the computer how it has been set up, chances are we can't tell the user either, because we have no formal rules and the user has to guess what the indexer has done.

BUCKLAND: Eliminating these terms with few postings to file under them is not the only possibility. One could just establish the connections internally and allow the user to query in his own language, which is likely to be specific, as Libby has suggested. With a console operation you can bring up these correspondences and allow users to work in a less constrained language.

TAUBE: There are all sorts of ways this could be done. I might add that when I say eliminate, we eliminate it from the printout of the index. We never eliminate it from the store of the computer, so that if later on a term becomes popular, we have had it in there, and at any time we can restore it to the indexing vocabulary.

ORNE: There is a tradition among librarians that has kept them trying to find a way to be consistent for some 50 years. The fact is there is a great deal of consistency in the library business despite the failure to reach accord on the cataloging code, or the filing code, or other things on which the British won't agree with the Americans.

It seems to me there is a fair amount of inconsistency within the machine profession. You have just gone back to the "unitermites," which I called them some 15 years ago. To me they're still "unitermites" and there are a lot of people that think the same about them. But they probably have their place, and they work for certain things. I only wish that in your language, in your own way of speaking, you could communicate with us as well as I think we are communicating to you what our problems are. We have tried many times to state what we want. You have tried this evening to put the resolution of the problem back into the hands of the librarians. You tell us, "This is your science." We know what we need to get out of it, but we don't know how. This is why I am here at this meeting, and it is the first meeting of this kind that I have attended in 10 years. This doesn't mean that I haven't read the literature. We are still at the same point, as far as I can see, that we were 10 to 15 years ago. We are not communicating any better.

TAUBE: I think you raised a serious point. This moves me to quote an epitaph I read some years ago that always struck me as very profound:

> Here I lie, bitten to death
> By the upperdog and the underdog,
> While trying to get in between them.

I well appreciate how the librarians have worked to develop consistent rules, and I certainly think that the only thing that is going to make the machines work is the consistent pattern of rules developed by the librarians. But that means that the librarian cannot stop; he must go on working on this thing. He can't say "I've gone so far, you take the ball; you do it better than I can do it." That is no answer.

EMLING: This is where I get bitten between the two dogs. I confess I know nothing about librarians or library science and not very much about computers, so this puts me in an excellent position to speak. I have the feeling that the first place for mechanization is in processing. Hopefully, the processing operations, which are complex, bookkeeping kinds of operations, the things people know how to do on computers, might well be carried out more economically mechanically than manually. If this is possible, then the bibliographical benefits might come to you largely for free. The mere fact that you could set up your files to do the processing would give you a nice start down the line toward bibliographical search. Now this is the way many of these things get started; there is an evolutionary process. You don't create a grand and glorious system and wait until you can realize it. You find some way of building into it.

This processing, if you're going to prove it in, is purely a matter of cost accounting. This seems to be something that is missing—your necessary cost data—but I don't think it is at all hopeless to get it. There is a way to work into mechanization, but there is information on cost accounting which all of you really ought to be working on to see if you can prove this thing in.

Now the other possible benefits from automation are in your bibliographical or search processes, and as I listened to this discussion I had some doubts. The question was raised by one of you gentlemen: What if we build our files on the pattern of the catalog that we have right now; what will the machine do for us that the catalog won't do? Now, it seems to me that it won't help a great deal if we only ask simple questions. If we propose simple questions and can get direct answers a manual search in the catalog is probably just about as good as a machine search and saves a lot of machinery. But there is one thing that you can do here: you can benefit if there are long and complex questions, because if you must "see also" this and "see also" that, you have a very long process if you go at this manually, and we all know that if it is long enough it becomes discouraging and people won't do it. You don't really know what the situation is, but you suspect that people don't use the library catalog as they should because it is just too much work. This is an area where the machine can do something for you because it can shorten this long and involved search process, but here again I find some missing links. You are asking the question: Should we have consoles or shouldn't we? I don't believe that is the question at all, because that obscures the definition of a console. It seems to me that if this long search process really is the reason we want a computer, then you must have some sort of a dialogue with the computer because otherwise you have to program

735-898 O—64——8

into the computer every contingency that anyone can ever think of. This doesn't sound very sensible. I think the real question is not whether we need consoles or not, but rather, to what extent is this complex search important to us? We ought to know to what extent we have these long searches. Then the question, as I said, is not: Do we want consoles or not?, but: How simple or complex should the console be? What is the reaction time that you need to compete with the manual search? These are the things that dictate the kind of console you need. I don't want to give you a lecture, but I just want to tell you what it looks like to an outsider.

WARHEIT: As a person who has had one foot in each camp, I'd like to say frankly that the librarian has had 50-odd years of experience, whereas, in the information retrieval business, the computer man has had less than 3 years of experience. He has had experience only with relatively small files, and he doesn't really know what he can do or what the machine can do. This is why I said this morning that the computer will do certain processing work for you. This the computer man knows; he knows the various clerical functions, and the computer manufacturer has been taught by the users what the applications are. It's been very interesting to me in the computer field to watch us pick the brains of the user to find out what we can do with the hardware. I agree that the beginning will be in various processing and clerical functions in the library Then the librarian will find out, on the basis of his total experience, what the computer really can do. We're speculating, at this point, and anything the computer man says is just second guessing.

TAUBE: A few years after computers were introduced, *Business Week* published an account of the use of computers in business, and 50 percent of them were admitted failures in business applications. I think the score is a little better now, but one of the things that the business community found out was that you couldn't take a computer, set it in the midst of the same type of organization that you had before, and expect to get a payoff from the computer. I have said before that the first automobile was a smelly toy and obviously not as good as the horse and wagon to get around with. There were no spare parts; there were no roads; tires were bad; it was a rich man's toy. But the automobile remade its environment. Now maybe this wasn't a good thing; maybe we'd rather go back to the horse and wagon, but as it remade its environment it began to pay off.

Now if the librarian says, "This is what I do; this is my pattern of service; put a computer in my library, and show me how it will do it better," the answer is, it will not. You might as well face that, it will not. Librarians have been smart enough to develop various instruments which do what they are designed to do fairly well. But are librarians willing to change the way they have been doing business because of new mechanical developments? Now again I've seen this in industry. We are consultants for a large chemical firm that has a lot of computers. This firm prided itself, before getting the computers, on the decentralization of its operations and on the fact that its various managers didn't talk to one another. Well, putting a computer in the center of this business didn't do any good; nobody would talk to it. Nobody would use a common language; nobody would give it codes; nobody would give it information, because then it might be available to the other manager. Well, this failed. In order to get the computer to pay off, this company had to reorganize the way it did business. Now if you say that the librarian is content, that regardless of what types of advances there are in technology that we stop with the typewriter and the 3 by 5 card, you might as well forget the computer.

PATRICK: In justifying a computer installation there are five things I look for. Anytime I can find a volume application which is well defined and repetitive, this is a candidate for automation. If it's well defined and repetitive and occurs in sufficient volume, I'll look at it at no cost to the client. I'll look at his cost, predict future cost, and we can justify it and go or not go on a strict cost basis. You all have such an application with these huge files, and this is why Libby and I both wrote about large files. You don't have this in a collection of 100,000 or less; in a file of over a million entries you do have such volume.

Another potential computer application is any application which requires sterile handling—where you must be absolutely sure of control of the information, control of the file, and can tolerate no errors. This is a computer application, and although the cost may be excessive, the quality is worth it. Most of our classified document files are on computers just for this reason. I can bond and security check 1, 2, or 3 individuals and I don't have to security check 500. All of our military security is this way.

A third way I recognize a computer application is by determining whether the application is so complex that a human being cannot do it properly. With computer techniques there is a hope, just a prayer mind you, that we can get it done right just once. If I can do this right just once, I'll code it, and I'll get somebody else to check my coding; I'll put this coding, this program if you will, in the computer, and then I can do it right every time. An example for you who are mathematically oriented is the solution of large sets of simultaneous linear equations, say on the order of 20 or more.

A fourth potential computer application concerns response time. If there is a strong payoff function-vs.-time, if the risks attendant to delay are significant, then you have a computer application. This is where the military money is going; we call this command and control. We've got about 30 minutes from the time the Soviet Union fires a ballistic missile at our country to do something, and we don't have 31 minutes, and we don't have 35 minutes; the penalty for delay is beyond comprehension. Consequently these are computer applications. You probably don't have any of these in librarianship because nobody is that interested about getting that book.

The fifth way, and this again has some meaning for librarians, is a situation requiring multiple hands in a file that must be current. This is something which, I'm sorry to say, you haven't brought up to me today, and it's my duty to bring it up to you. If you publish a book catalog once every 6 months, that dog is a 180 days out of currency—old, if you will—on the last day you use it. We are all aware of the delays in publication. In the scientific and aerospace fields, where I work quite frequently, that kind of delay is absolutely intolerable. Consequently we use computers to keep these files; we can make many selections from the same file in a very short period and the file is always current. This is the main reason for putting the bibliographic file on the computer and interrogating it with consoles. When you interrogate it with a console, you are interrogating the *current* holdings of the library.

On the airplane coming in, just for the fun of it, I figured what it would take to update the National Union Catalog if I had it on magnetic tape—current technology—the kind of hardware we can order today from IBM, CDC, RCA, and so on. It would take 10 hours. Now that's a long time in my parlance—it is nothing to you. It would take only 10 hours to process all the daily reports against the National Union Catalog and have a current copy of the National Union Catalog with 14 million entries in it. Therefore, every night after most of the people go home the day's acquisitions could be processed and when everybody comes to the library the following morning the catalog would be current. Now, I don't know what this is worth to you, but you've never experienced it before.

Computer technology is only 13 years old. In that 13 years computers have increased in speed by about 4 orders of magnitude: that's 10,000 times in speed. Computers have been reduced in price almost 2 orders of magnitude in that same period. So our cost-per-dollar ratio is 10^6 times what it was when I first got into the field. I hate to overstate the case, but it's almost a revolution in technology. We can do things now for pennies that it used to take a research librarian months to do. I think you should be aware of these things.

TAUBE: I made the point earlier that the librarian has depended for much of his bibliographical service not only on the card catalog but on the abstract services and on the indexes sponsored by the scientific societies and other organizations concerned with processing bibliographical information. Now it may turn out that these organizations are going computer before the librarian does, and these organizations will not print anymore. They will not print the decennial indexes, and they will not print the critical tables, because of the volume of material involved and the paucity of use. The librarian who wishes to serve his customers beyond the monograph may have to key in with computer systems which analyze the important scientific and humanistic literature of his day.

WARHEIT: To illustrate your point, Dr. Taube, a couple of weeks ago the medical librarians on the West Coast had a meeting, and I never met a more sober group of librarians in my life; they suddenly realized that next year they were going to get material on tape from the National Library of Medicine. They have to use this material; there is no question about it.

DIX: It seems to me that our concern here is, and ought to be, what can we do to solve simple problems. I think some of us here need to know more about this console and how it would respond to the simple and most common kind of library search: Do you have this book?—and the question following from that: If so, where is it? Now how is this question asked in terms of simple manual operations? Does the scholar have to come in and try to type something out on a typewriter? How is he going to feed this in? What is going to come back to him? Can he do this in the same amount of time in which he can walk to a tray of cards and leaf through them or pick up a big book and run down the column? Is it going to cost as little? Now I know one answer, of course, is that we don't take this aspect alone because there are a lot of other byproducts you can get here, and I respect this. But, one by one, I wish we could tackle these very specific and simple daily library operations; we need to know what the state of the art is in this kind of thing. In other words, is there a machine now on the market that will take this store of knowledge and enable someone to query a specific library and get a specific answer at the same time that, let's say, 50 or 100 other people are asking the same kind of question?

F. B. ROGERS: It seems to me that that kind of question gets you nowhere. The trouble with librarians is that they say all we want is everything. You will never get anything at all in this way. What the librarian must do is decide what his purposes are and he cannot state these purposes in large general terms such as: What I want to do is to make my operation more efficient. He cannot state these in very specific terms, such as: What I want to do is to get a set of cards for one-half cent a card. This absolutely will not do and will get you nowhere. You must decide what your major purposes are, stated in some terminology that people can understand. You must not only state those purposes, but you must rank those purposes in some order; you cannot act as if everything you do has an equal value on this scale. I would say that when the librarian does this, the real value in getting together with machine people is to get some idea from their talk of what kind of new constraints operate in their area and, by doing so, begin to get some idea of the constraints under which we now operate and of which we're largely unaware. You just get some ideas this way. But you must tell them what your major purpose is, and what you want to do and not just say: Can I get a console that I can ask, "Is this book in the library?" There is absolutely no percentage in that, and we will get nowhere if that is the direction in which we go.

LIBBY: Dix has asked two questions or a two-part question. The first concerned the mechanization of the control of processing of bibliographical items within a library with respect to getting up-to-date and quick answers as to whether an item is on-shelf or where it is if it is not on-shelf. Another aspect is the storing of serial titles and getting automatic lists as to whether they're overdue, and the preparation of bills or payments with respect to acquisitions or services rendered by the library. I believe that present off-the-shelf computer equipment can solve these problems. I believe that they can be solved on an economic competitive basis with present manual techniques and with better performance than is now achieved. I would suspect that in larger libraries there is great uncertainty as to what has been received, whether items should be paid for, and so forth. The present state of computer technology can handle this type of operation for a library.

Now the second part of the question, or my interpretation of the question, concerns the mechanization of the bibliographic descriptive data and the servicing of it to the user in automated libraries. There is no clear cut answer as to whether this is worthwhile on a dollar-and-cent basis. It is going to have to be, I feel, a decision on the part of the librarians as to whether the following type of service is valuable, and I will try to answer your question by describing how one might envision a user operation.

Is it worthwhile (this is going to be difficult to answer) for a user to walk up to a device, and, if he knows an author's name, press a little button marked "Author," and then with one finger, possibly two, type in the last name or the last name and initials? He types in, say, "Smith" and gets an immediate response: "There are over 3,000 catalog entries pertaining to this name. Press the

Authority-Catalog button and enter the name again." He does this and gets some instructional material about looking under Smith, Schmit, or is asked to put down some subject term as well as the author's name.

Suppose he starts by pressing a button that says "Subject" (I'm not using subject in the sense that you librarians use it), and he enters words that he thinks pertain to the subject about which he seeks information. He gets a similar indication immediately from the device that 500 or 50 or 10 entries relate to the set of terms that he has entered or, if there is no response from the catalog, it displays the relevant segment of the hierarchical subject categories and asks him to select terms which he feels are closest to his needs.

Now it is beyond the technologist to decide what this kind of service is worth in dollars and cents; I think the librarians have to answer these questions themselves, not from an administrative viewpoint but strictly from a user-service point of view.

WARHEIT: I want to illustrate something on the point Dix raised. I was in our Human Factors Laboratory watching people work with the console. They were doing a rather simple operation of writing invoices. The clerk had an order and would go through a series of catalogs manually to determine the code number, quantity, price, discount, delivery date, and a few other things like that. That manual operation was then compared with the operation on the console where the clerk pushed the button for the desired portion of the "catalog," and recorded from it. The accuracy, speed, and throughput were much higher and the error rate was much lower than in the manual system. It was a complete analogy with a person walking up to a manual card catalog and going to various trays and extracting information.

FUSSLER: It seems to me that in response to the question of priorities, it may be possible to isolate some of the issues and thus benefit from the present company and further discussion. I would take exception to some of the implications from Jerry Orne, if he infers complacency with the existing system in large research libraries over any prolonged period of time into the future. I think the system is deficient in a number of critical respects as it relates both to readers and to internal processing operations, to the extent that these can be separated. Now this is an opinion and it is quite obvious that there may be different views on it. It is hard to adduce much evidence on the degree of dissatisfaction of readers at the present time, since studies with useful data on the performance of readers with respect to bibliographical apparatus of large research libraries seem to me singularly deficient and inadequate. There is, however, a good deal of evidence with respect to internal processing or operating difficulties. There are problems when the reader is trying to deal with card catalogs with 3 or 4 million cards and up.

For these reasons, and some connected with them that I have not stated, it seems to me that we are really obliged to move and move as rapidly as possible to alternative means of handling these problems, using as criteria either costs or benefits and preferably both wherever it is possible to do so. This would seem to me to suggest that the issues for this group are how and where to start, avoiding the two extremes that have been discussed today. I move that we don't wait for perfection with respect to traditional library processes because we will never have it, and that we don't wait for the perfect system of automation because we are unlikely to have that.

As to the second issue, it would seem that before we start we should try, with the most sophisticated advice that we can secure, to define the characteristics of the long-term, basic, mechanized system that is most likely to emerge, the requirements that it would impose in terms of standard operational procedures, and so forth.

I'd like to add a footnote. This is a matter of personal choice in terms of priority, but I think I am speaking for some librarians when I say that the internal processing job needs to be cleaned up before we can get into extensive expansions of benefits for readers. Generally speaking, librarians have moved hard in terms of applying available resources to increasing reader benefits.

DIX: Brad, would you define a little more the kind of questions we ought to be asking? If you think that I'm asking the wrong question, and I realize that it is a very elementary question, we might spend a little time talking about what kinds of information each group here needs from the other.

F. B. ROGERS: What I meant to indicate basically is that surely the question is not: How can I automate my library? This is a meaningless question operationally at this hour in history. First we have to identify as best we can some of the crucial areas of our operation. We have then to study carefully the size of the different factors which enter into those situations. It is hard to give an example because this is the most difficult job in the whole analysis. At my library when we began thinking about this some years ago, we did put down on paper, with great agony, what the problem was that we were attacking and exactly what it was that we wanted to do. Today those things seem so simple and so obvious to me that I don't see what other answers were possible to us. I wonder how we could have struggled as hard as we did to find these objectives. But it was a struggle. I think, having identified these crucial areas, we are going to have to make up our minds realistically to some of the prices that we are willing to pay.

I have accepted and still do accept as a truism the idea that it is efficient to compile bibliographic records centrally. All of my concepts are built along this line. If you accept this simple idea you have to go another step and do some serious thinking about the standardization you will accept. Librarians just won't accept this kind of standardization. To almost everything that is proposed, some wise man from the East gets up and says, "But how about this instance? It won't cover this instance!" Now we have to give up one thing or the other. We cannot continue this kind of argument about standardization and be unwilling to accept standardization and talk seriously, at the same time, about central compilation of the bibliographic record. They're just contradictions and it won't go. It seems to me that this is the sort of thing that we have to begin to try to realize and to act on. To me this idea of summoning up the section on the "Smiths" on a console is such a trivial purpose that I would not spend 5 minutes trying to figure out how to do it or how much it would cost or anything else about it. Now I could be wrong about this; I said this is the way I look at it. I think my advice would be to look for some other area that is really important to do, and let's not waste our time with this sort of thing. At least that's a proposition for a debate, and it might illuminate the situation somewhat.

HAYES: I am consoled, if you will excuse the expression, by the direction taken by Dix's very fine question, which I feel is the appropriate direction for any such meeting as this, and I think that Rogers' answer is excellent. I would like to comment on it without trying to enumerate the set of questions that the computer people might have set down as those they want to ask the librarians; essentially they fall into three categories: cost, time, and function. Cost data are at best difficult to get, and they seem to be extremely difficult in the library field. I suppose it is because we are dealing with a very complex intellectual process. As a step in the direction of trying to answer it, I gave a course at UCLA in which I had a group of librarians and a group of people from the School of Business Administration. I assigned as a problem the development of a cost accounting system. The students started out, as librarians normally do, by dividing it into functions, which is a very reasonable way of cost accounting. But there is another way of cost accounting which I then suggested that they pursue—process accounting, in which you define, not functions that you are performing, such as circulation, reference, and the like, but the different types of processes which are involved. This is a much harder thing to cost out, but it is the type of thing that the computer people want to know; they want answers to such questions as these: What classes of questions are we going to handle? How rapidly do questions have to be responded to? How much are we going to pay for them?

The effect of the computer has been mentioned, and I think that the principal effect is that of clarifying these questions. The effect very frequently is that if you institute the changes or clean up the processes, the computer is no longer required. The program to carry out the intellectual part results from the defining of these processes that are involved in library work. As to what role the computer can play, I agree completely with Brad Rogers—at least this was the implication which I got—that the procedural aspects can very quickly and reasonably be placed in a computer. How much of the intellectual aspects can be, is debatable. Why the computer people want to investigate the intellectual aspect is another matter; I suspect it is because it is interesting and difficult.

ELLSWORTH: Could I ask a question of Libby and Patrick—In writing your reports, were you able to lay hands on written statements by librarians as to what we are doing and what our needs are?

LIBBY: I personally have not received a good picture as to whether librarians are more interested in a smooth-running operation or in potential improvements in service to the library user. I have a predominant impression, from this meeting and others, that librarians are interested in the potentialities of mechanization primarily from the point of view of making their job easier, rather than serving and increasing the use of the library; now that's a personal impression.

ELLSWORTH: I strongly suspect that the literature that we have written would sustain this point of view, but I'd like to have some other views.

PATRICK: I did a reasonable literature search at the UCLA Library. I found only two documents useful to me, one was the Schultheiss work that came out of the Chicago study[17] and the other was Fussler's report on book use.[18]

TAUBE: I might tell a little story which indicates that, even though the computers take over, there will always be work for those who are interested in manual catalogs. We got a call the other day from a railroad company official who said, "We understand you people have developed a manual indexing system." I said, "Yes, we have developed such systems; what is your problem?" He said, "We have so many computer tapes that we need an index!"

Now to show you that everything comes in twos, I visited the Social Security Administration the other day; they have 38,000 computer tapes, with a manual library system for indexing and cataloging those computer tapes.

[17] See item 67, p. 139.
[18] See item 3, p. 88.

CLAPP: A little note of acerbity has crept into this session, and perhaps it is just as well, because this indicates the anxiety, enthusiasm, and frustration which exists on all sides. It doesn't do any good to say I'm never going to come to another one of your meetings unless you produce something at this one! It doesn't do any good to threaten unless you do something right away, I'm going to bury you! This is the whole point of this meeting. Ladies and gentlemen, if the people who wrote the state-of-the-art papers had known what the participants were going to say, they would either have not written the state-of-the-art papers or they would have written them entirely differently. The whole point of the state-of-the-art papers is to excite you.

A while back you applauded Emling, who, as I gather, was pouring oil on waters and doing a very nice job of it. Nobody has picked up his point which was that the introduction of a computer technique into library work is to take care of the complex question. I don't think we've discussed this adequately, and it's probable that we won't discuss any of the working papers really adequately. It is certain that we're not going to walk out of here with answers, either from the technician's or librarian's viewpoint, but I think it is equally certain that we will all walk out of here with an improved understanding of the whole situation. Brad Rogers speaks with the certitude of hindsight with respect to a very important operation. Ten years from now we may all speak with equal certitude of hindsight. But meanwhile until that 10 years is past, Bill Dix's ruminations, and perambulations, and projections have just as much validity as Brad Rogers' agonizing 10 years ago. Only hindsight will prove this out. With all due respect to the utility of acerbity in discussion let us now adjourn.

SECTION IV

Graphic Storage

The Current Status of Graphic Storage Techniques: Their Potential Application to Library Mechanization

SAMUEL N. ALEXANDER, F. CLAY ROSE
Data Processing Systems Division, National Bureau of Standards

Introduction

The purpose of this report is to describe briefly some of the newer technologies, particularly in the field of microforms of graphic records, and to consider their probable effects on the operating procedures of libraries. Graphic storage deals with the essential materials with which the library works. Today these materials are predominantly in the form of bound volumes of the printed page and the typed or printed catalog card. The problem of providing effective service in the face of a steadily expanding volume and variety of literature is becoming increasingly critical. Thus, it is both timely and wise to assess the cost and space alternatives afforded by the new storage methods, especially those involving miniaturized facsimiles of the library's materials.

While at present the compelling motivation may be that of savings in space and cost, microform facsimiles, together with means for their mechanized selection and retrieval, offer more than an alternate approach to "housekeeping" problems. For example, deterioration and loss of rare and irreplaceable materials are minimized through the utilization of associated techniques for providing copies. Further, this technology offers a potential means for attaining far wider coverage of the world's documentary resources through acquisition of materials reasonably available only in microform.

Before entering into the characteristics and implications of automated microform systems, a few clarifying remarks might be in order. Some librarians (and their patrons) have an understandable aversion to the usual facsimile reproductions because much of the esthetic quality associated with well-executed books and publications is lost in this process. Facsimile copies do not readily retain the charm of artistically illustrated materials or the attractiveness of finely detailed maps. The impressiveness of elegant bindings and the satisfying feel of a massive tome are gone. While this sense of loss may be meaningful in the case of rare books and valuable manuscripts, for much of the product of the modern printing press this can hardly be a major consideration.

There is a natural concern that the library's basic mission of adequately serving its patrons should not be diluted by introducing new technology. Certainly one would insist that, while seeking to live with its growing number of operational problems, the library must continue to serve its patrons at least as well as it does now. No doubt, as the newer technology is introduced, there will need to be some adjustments both in the working procedures of the librarian and in the inquiry protocol by which the patrons express their needs. The introduction of such technology may hasten the day when the librarian will no longer need personally to mediate in a large fraction of the routine transactions by which the patrons are placed in juxtaposition with the desired part of the library's collection.

Despite potentially attractive accomplishments to date, there is need for considerable refinement in the available technology. Moreover, there is even greater need for the evolution of the associated library procedures so that this technology can be applied effectively. Experience has indicated strongly that introducing technology without adequate prior planning and adjustment of the affected procedures often reduces the effectiveness of the technology to such an extent as to impugn the advisability of having acquired the newer equipment.

In an effort to show how these broad considerations apply to the specific subject of graphics, this report is presented in the form of a state-of-the-art precis that has the following sequence: First, a brief history of microform and related facsimile recording systems is given. Next, the nature of the intellectual task associated with designing an effective system is considered. (See item 90.[19]) These topics are followed by a discussion of facsimile storage and retrieval systems in terms both of their utility characteristics and of replication methods and media related to microform systems. The implications of this technology for the library environment are then examined, and finally conclusions that may be inferred regarding the direction that further technological development might take are presented. A set of system descriptions specifying their individual characteristics and a selected list of references are appended to provide the reader with access to such additional detail as may be of interest.

Brief History of Pertinent Developments

The advent of mechanization in the selection of graphic materials from storage antedates this conference by a little less than four decades. Even earlier, the possibilities for miniaturization of documentary materials had been developed as an extension of the art of photography. Particularly since the 1930's, there has been increasing interest in the development of retrieval devices that utilize the following principles:

1. A miniaturized or compressed form of document storage.
2. Means for the mechanized manipulation of stored microimages in the operations of "finding and fetching."
3. Means for display and replication of images selectively taken from the store so that they may be viewed and used by the "customer."

In addition, some of the earliest proposals, as well as many present-day devices, utilized the principle of integral indexing—that is, the inscribing of identifying labels or "retrieval hooks" directly on or physically adjacent to the items that are to be retrieved from storage.

The first principle, that of compressed storage of document images, dates back to the year 1839 when John Benjamin Dancer of Manchester, England, first combined the techniques of photography and microscopy to produce the microphotograph of a document. (See item 54.) The reverse of this process, the enlargement of the reduced image to provide a replica of the full-size original, is necessary to meet many of the practical requirements of human viewing and using.

The second principle, that of mechanized manipulation of records in a file, was realized by Herman Hollerith and others from about 1890 onward, through the invention and use of punched card and "needle-selection" card techniques. In recent years both of these techniques have been combined with microphotographic techniques, one embodiment being the microfilm aperture card.

The principle of integral indexing has recently received considerable attention. Surprisingly, its inception predates even the use of papyrus and paper for graphic storage. A physically integral index has existed at least since the time of ancient Sumaria, where it was frequently the practice to put a thin layer of clay over a tablet that had already been inscribed with cuneiform characters. This expendable layer would then have inscribed on it indexing clues to the information on the tablet itself. This principle evolved through the centuries, and one of its many manifestations is the present practice of stamping library classification codes on the spines of books and onto other material in a document collection. The modern counterpart appears in such mechanized systems as the Rapid Selector, Minicard, and Filmorex.

In the other approach the index to the mechanized store is a separate file or list of document identifiers, such as subject headings, descriptors, or classification codes, that lead the searcher to the "locators" or "addresses" of items in the store. This separate index may or may not be independently mechanized. One example of this approach is the original Recordak Lodestar, which is a mechanized microfilm retrieval display device that is manually set after reference to an entirely separate index. An example of another form of this approach is seen in the Microcite system where

[19] This and similar references refer to items in the bibliography on page 136.

an otherwise separate mechanized index in the form of "peek-a-boo" cards is physically coupled to activate the "retrieval for display" device.

Compressed graphic storage of books and records, with microphotographic duplication, probably first received practical application when the Department of Agriculture Library (now the National Agricultural Library) inaugurated the so-called Bibliofilm Service in 1934. At approximately the same time, in various parts of the world, serious consideration began to be given to the development of equipment which combined microform storage and retrieval of replicas through the use of integral indexing and mechanized selection. Patents issued to Goldberg in Germany (1931), Bryce (1939), and Loughridge and Stuart (1940) disclose various possible applications of these combined techniques. Even more significantly, from the 1930's onward, both documentalists and engineers, such as Atherton Seidell, Watson Davis, and Vannevar Bush, began to work toward combined techniques specifically applicable to library services. By 1940, Bush had developed the prototype Microfilm Rapid Selector, whose lineal descendants are among the systems currently available.

Today there are devices and systems of two basic types. The "address" system is, as previously noted, one that stores only the document images; the user must approach it with information derived from a separate source that identifies the specific document he wishes to see. The "search" system, on the other hand, combines a mechanized index component and the means for retrieval of selected documents from the store (see fig. 14). Both types, which are discussed later in more detail, can play important roles in the automation of libraries. As of now, however, there has been little practical application in the conventional library environment.

The Importance of the Systems Problem

It must be recognized that hard-core intellectual problems underlie and are inherent in the library situation. (See items 7, 18, 30, and 66.) These problems are, in the main, independent of mechanization. Honest differences of opinion in categorizing document content as to meaning and relevance persist even among specialists in well-defined subject fields. This indicates that the

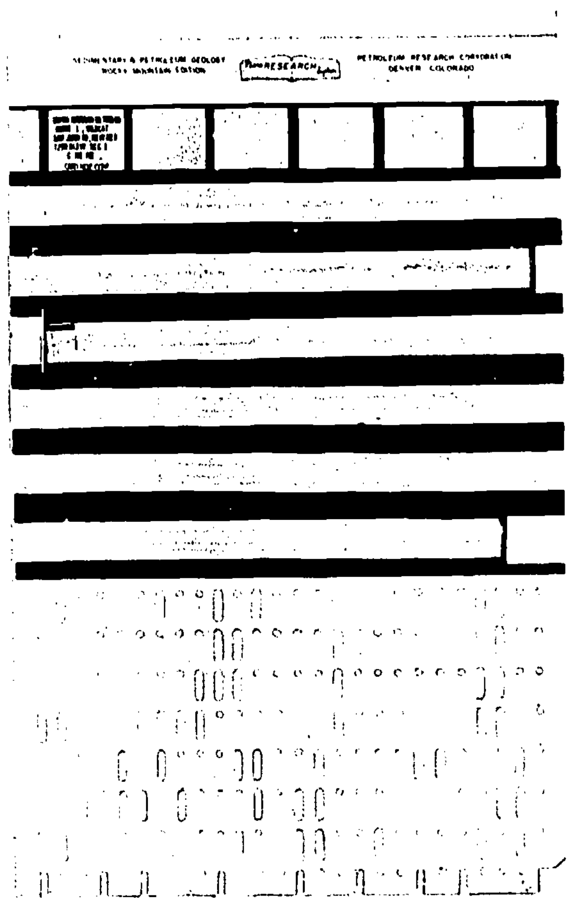


FIGURE 14. *A "search" system—The Micro Research System, unit microfiche and needle sort slots.*

"heart-of-the-matter" problems will be with us in the years to come, and that they can neither be solved nor dissolved by the conveniences offered by new equipment or by esoteric classification and coding schemes. However, there are intellectual problems of a somewhat more tractable nature that are posed specifically by the availability of new technology.

These intellectual problems involve balancing the alternatives presented by different systems. One must always give up something to get something else; for example, one system might have quick access but low resolution, another system might allow one to move images around and enlarge parts of them but it might be restricted in capacity. The problem of alternatives then is the essence of systems design and such problems can yield to a well-planned attack.

In seeking to apply new technology to the needs of the library community, there is often insufficient attention given to the practical requirements of the user. Thus far, the library community has not followed the lead of the business world in rushing into the use of this new equipment without recognizing the intellectual requirements basic to its successful application. In general, business has not utilized microform to its full potential. Whether librarians can avoid a similar fate is questionable, because recent high-level agitation has increased the danger that the carefully planned systems approach may lose out to a "reductio ad gadgetum" attitude. Thorough systems planning, in the library as elsewhere, means the development of an effective and economical balance of man-machine efforts within the total system.

Specifically, it should be recognized that the difficulties of mechanizing library procedures relate far more to decisions involving document analysis, subject-content indexing, and machine coding than they do to the characteristics of either the equipment or the storage media. It does not gain much to put documents into a miniaturized storage system if the method for retrieving them will be no more effective than what we already can do. The way in which specific user-oriented requirements and the man-machine capabilities are fitted together into an integrated system will determine the success or the failure of a particular mechanized technique. We cannot hope to escape from the human factors involved in analyzing the subject content, on the one hand, and in evaluating and using the products of mechanized search and retrieval, on the other.

Perhaps the librarians have sensed that their major problems cannot be solved merely by the installation of equipment. This may be the reason why relatively little utilization of mechanized graphic storage and retrieval systems has been made to date in general libraries. Nevertheless, new tools are finding useful applications in a number of specific situations. Knowledge of what is available should help to direct and motivate the prerequisite systems planning needed for their eventual successful application. Familiarity with performance characteristics of available equipment will be required in order to make sensible decisions in determining the appropriate levels of content analysis, in choosing coding systems, and in providing for open endedness so that the system can be adapted to the changing conditions of actual use.

Systems Characteristics, Media, and Replication Methods

The present state of the art in mechanized graphic storage may best be appraised in terms of the performance and other characteristics of a variety of devices, storage media, and complete systems. (See item 47.) Details of the characteristics of each system that is actually or potentially available for library use are tabulated in appendix A based on data reported on the form shown in figure 15. Pertinent information on many systems is still not available (see footnote 1, p. 134).

As shown in figure 15, each system or device is first identified by its name and by the name of the developer or manufacturer.[20] Next, availability status is shown. A system or component is reported as operational and commercially available only if it is currently offered on the open market for a more or less determinate dollar cost. Otherwise, a system may be (1) operational, but not generally available; (2) developmental, that is, either the entire system or certain of its components are in various stages between design study and testing; or (3) existing only as a formal proposal, although various feasibility studies may have been carried out.

Identification of systems as to functional type is based on the distinction between search and address approaches previously discussed. That is, the address system, which contains documentary material in some form of microform storage, is approached by a searcher who has obtained the necessary locating information from a separate index.[21] This system has the capability of displaying or reproducing material for which the

[20] See items 2, 5, 9, 11, 12, 13, 19, 29, 34, 39, 40, 58, 65, 69, 70, 71, and 76 in the bibliography, p. 136.

[21] It should be noted, of course, that possibilities of microform storage and machine retrieval exist for index systems proper and that, in fact, the mechanization of the card catalog may be one of the most intriguing future applications of this technology in large libraries.

SYSTEM DESCRIPTION

NAME: ____________________

DEVELOPER/MANUFACTURER: ____________________

STATUS: ☐ Operational ☐ Commercial ☐ Non-Commercial ☐ Developmental ☐ Proposal

TYPE: ☐ SEARCH ☐ ADDRESS

SIZE: ☐ Small (<$10,000) ☐ Medium ($10,000–$200,000) ☐ Large (>$200,000)

PURPOSE: ☐ General ☐ Special ____________________

TIME FUNCTION: ☐ Immediate Response ☐ Delayed Response

INTEGRATION FUNCTION: ☐ Off-Line ☐ Shunt ☐ On-Line

INPUT SIZE: ____________________

STORAGE MEDIA:

- ☐ Transparency/Translucency
  - ☐ Microfilm
    - ☐ Roll
    - ☐ Strip
    - ☐ Scroll
  - ☐ Microfiche
    - ☐ Unit
    - ☐ Chip
    - ☐ Sheet
    - ☐ Jacket
    - ☐ Slide
    - ☐ Aperture
- ☐ Opaque
  - ☐ Microcard
  - ☐ Microlex
  - ☐ Microprint
  - ☐ Microtape
  - ☐ Electrostatic Print
- ☐ Electronic
  - ☐ Video Tape
  - ☐ Other
  - ____________

STORAGE CODING: ____________________

STORAGE UNIT CAPACITY: ______________ STORAGE DENSITY: ______________

SELECTION: ☐ Automatic ☐ Magazine? ☐ Semi-Automatic ☐ Magazine? ☐ Manual

AVERAGE ACCESS TIME: ____________________

OUTPUT: ☐ Display ☐ Copy ____________________

PRINTOUT TIME: ____________________

SYSTEM FLEXIBILITY: Yes ☐ Update/Change No ☐ Yes ☐ Add/Purge No ☐

FIGURE 15. *Sample data sheet for graphic storage system description.*

searcher specifies the address. A search system combines both index and address approaches by including a mechanically searchable index that is functionally, and in many cases physically, integral with the microform copies of the desired documents.

The next item of figure 15 is the size of the system, as crudely measured by its general price range. Systems offered for less than $10,000 are designated as small; those ranging from $10,000 to $200,000 as medium; and those costing more than $200,000 as large. Other significant ways by which the size of a system could be indicated are by the document storage capacity, the physical space required, and so on. However, cost probably provides a composite indicator that is sufficient for our purpose here.

The response time of these systems and their functional relationship to automatic data processing systems are also indicated. While a document library is not normally required to provide an immediate response to a request, some storage and retrieval systems, particularly those designed for command and control and/or intelligence data displays, do try for immediate response. An example of an immediate response system is ARTOC, described in appendix A. The trade-off for this response capability usually, as with ARTOC, is a limited storage capacity. These on-line systems usually are specialized ones that can achieve decisional responses immediately following the selected presentation of current data. Thus far, they have been applied mostly to the control of valuable or "perishable" inventory or to military situations.

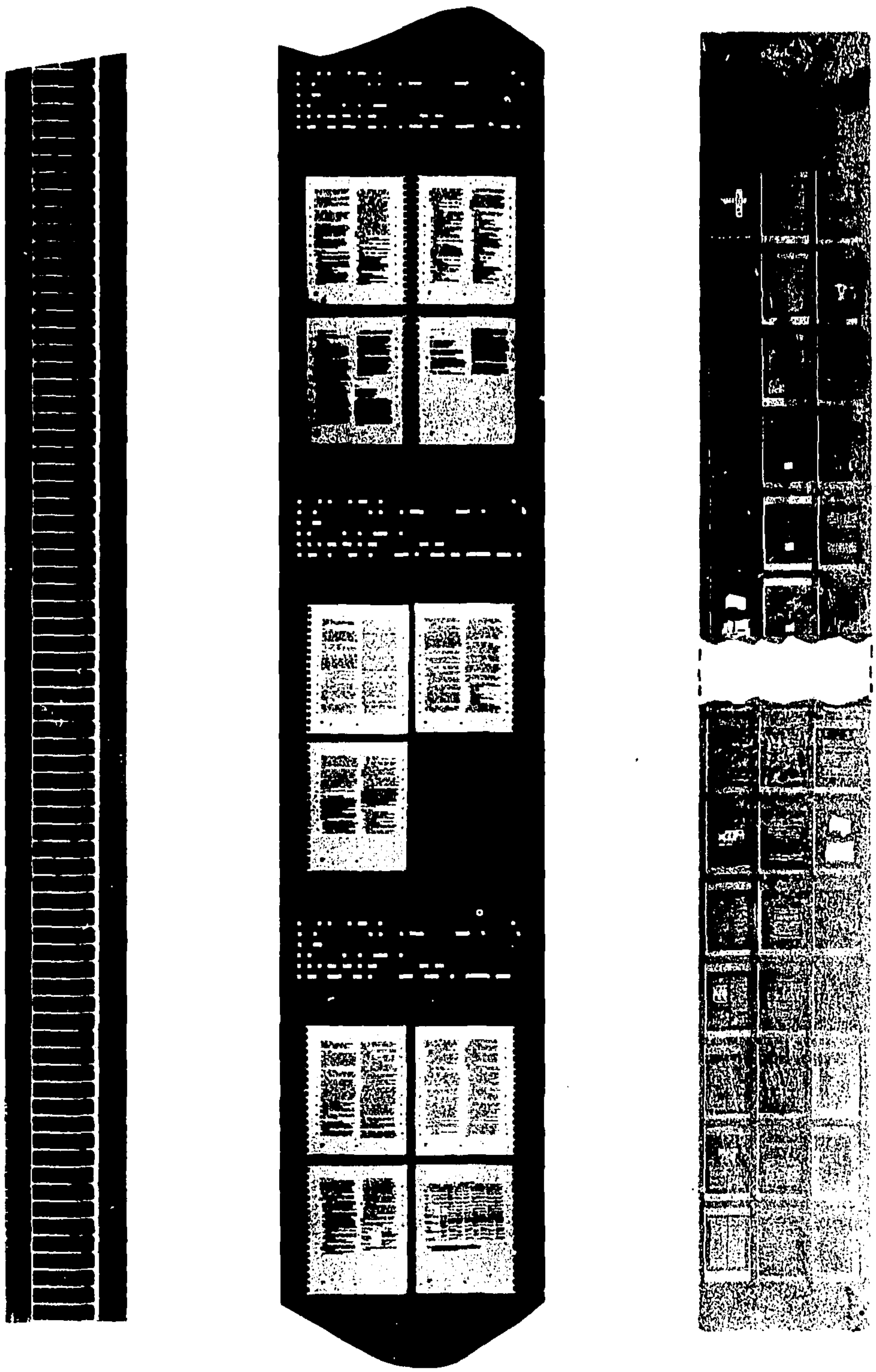
These on-line systems tend to have a functional or actual tie-in with the associated data processing facilities, and this arrangement is not too different from what would be an effective one for many conventional library situations. However, most of the systems that we will discuss are being employed as off-line systems, in that they have no direct tie-in with the data processing facilities. To emphasize this point, we designate as "shunt" systems either those that use ADP facilities on a part-time basis for industry research, or have a computer as an integral part, or operate as terminal equipment that can be used either on- or off-line with an ADP facility. For the most part, these shunt systems are currently expensive configurations that call for extensive search requirements in order to justify their present costs.

Consider now the input to a system. For our purposes, input is generally restricted to textual and pictorial information that will be absorbed in toto by conversion to microform. There are limitations on the size of input depending on the photographic or television camera used and its resolution capability. These factors influence the choice of a suitable storage medium. For example, the current trend is to reproduce textual material onto media the size of 16 mm film and maps, charts, engineering drawings, and similar material onto 35 mm media. This practice is based on conventional microfilming techniques and does not consider some of the more recent types of conversion such as to video magnetic tape, thermoplastic media, or photochromic storage.

Images in microform (see item 38) may be stored on any one of several media. Storage media include transparent/translucent film, opaque film, paper, card stock, and magnetic video tape. The transparent/translucent microform may be a fixed microfilm roll (16, 35, or 70 mm wide and many feet long). Document pages are sequentially arranged on this roll and each page covers the usable film width. (See items 23, 63, and 64.) Alternatively, it may be strip, which is made by cutting a roll into specified lengths; or scroll, which is 10 to 20 inches in width and several feet in length (as for the CRIS system described in appendix A), and on which many page images may be placed across the scroll width. (See fig. 16.)

The transparent/translucent microform may also be in the form of a discrete unit (i.e. one having an alterable sequence), such as a film chip, transparent plate record, or microfiche. (See items 8 and 88.) The microfiche occurs in several different varieties, the most common of which are described and illustrated below:

1. The unit record, a piece of film, usually no larger than 5 by 8 inches, on which a few images are recorded. A variety of unit microfiche records is illustrated in figure 17.
2. The jacketed microfiche (fig. 18), a record approximately the same size as the unit microfiche. It is made up by inserting microfilm strips into individual sleeves of



a b c

FIGURE 16. *Transparent/translucent microfilm. a. 16 mm roll microfilm, NBS FOSDIC II. b. 35 mm roll microfilm, NBS Rapid Selector. c. Strip microfilm, IBM Walnut.*

735-898 O—64——9

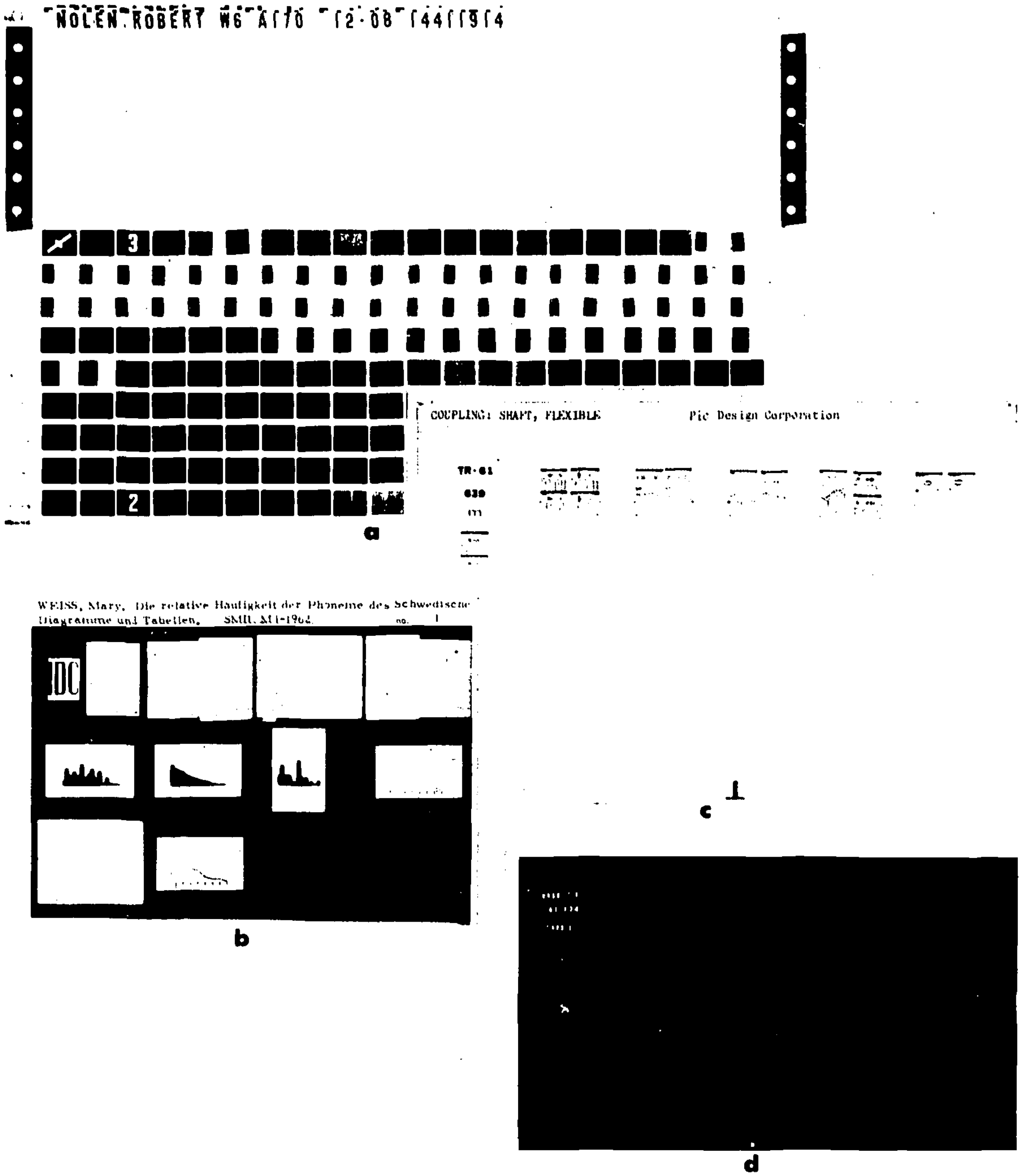


FIGURE 17. *Unit microfiche. a. Recordak Corp. b. International Documentation Center. c. Thomas' Register d. Microcard Corp.*

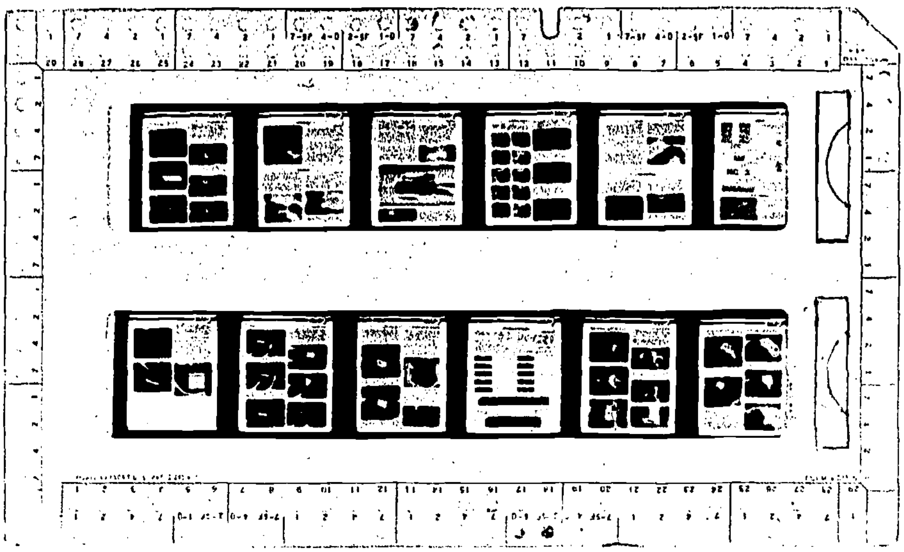


FIGURE 18. *Jacket microfiche—McBee Keysort Card for 35 mm (includes needle sort notches).*

a transparent acetate jacket. Several such jackets are shown in figure 19.

3. The sheet microfiche, somewhat larger than the unit microfiche, on which several unit records are usually recorded (fig. 20).
4. The aperture card, a card into which a square or rectangular hole is cut and a chip of microfilm mounted. It may be an index card, a machineable electronic accounting machine card, or an edge-notched mechanically sorted card (fig. 21).
5. The slide (fig. 22), a single microfilm image mounted in a frame for ease of handling although groups of slides are sometimes magazine-loaded into a system.
6. A chip microfiche, a discrete unit, usually containing a single image or a small number of images and so small that it is normally manipulated with others in a cartridge or magazine, or on the "shish kebob" skewers of a Minicard system. Several types of chip microfiche are shown in figure 23.

The opaque microform, another variety of discrete unit record, presently takes one of four forms. (See item 72.) The most well-known is the microcard, which is usually 3 by 5 inches, has an emulsion

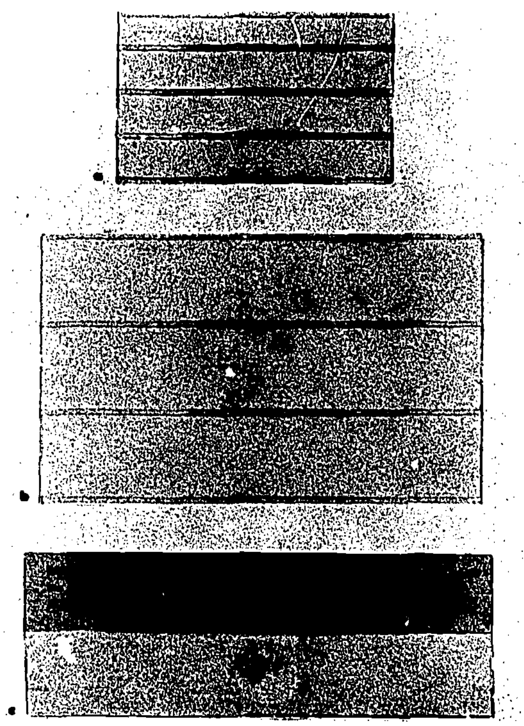


FIGURE 19. *Microfiche jackets. a. NB Jackets Corp. Microjacket acetate—16 mm. b. NB Jackets Corp. Microjacket acetate—35 mm. c. Sertafilm, Inc., acetate—35 mm.*

onto which document images are photographed, and is most commonly seen as the product of the Microcard Corp. Microtape, developed by Microtape Systems, New Haven, Conn., uses a photo emulsion on a paper stock which is backed by a pressure-sensitive adhesive. Its normal width is either 16 or 35 mm, and it may be several hundred feet long. The usual application is to cut the tape into strips which are then affixed to cardstock. Microlex, of the Microlex Corp., Rochester, N.Y., is an opaque film sheet, approximately 6½ by 8 inches, on both sides of which document images are recorded. Microprint, produced by Readex Microprint Corp., is an opaque sheet of paper, 6 by 9 inches, on which microimages are printed by an offset process. Figure 24 illustrates typical opaque microforms.

Prospects for a form of facsimile storage that produces electronic signals directly as output are currently represented by the use of video magnetic tape, usually 2 inches in width. Such "electronic" storage provides for facsimile recovery of full text and pictorial material. Another related use of magnetic tape recording, from ¼ to 1 inch in width, is that of digital storage in coded form of short descriptive text such as accession number, descriptors, or possibly a bibliographic citation. (See item 74.)

Present systems are generally restricted in their ability to produce other than black-and-white

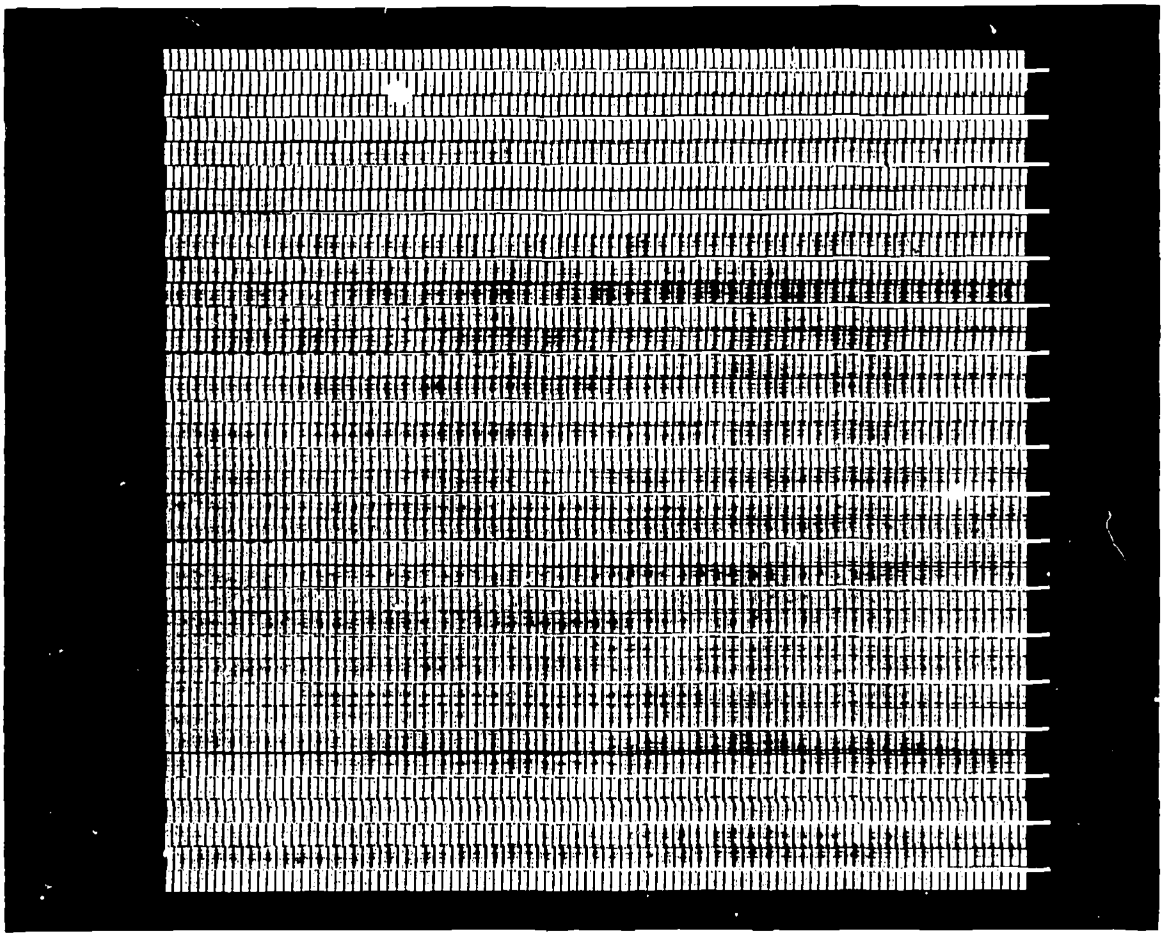


FIGURE 20. *Sheet microfiche—National Bureau of Standards "Microcite II."*

a
b
c

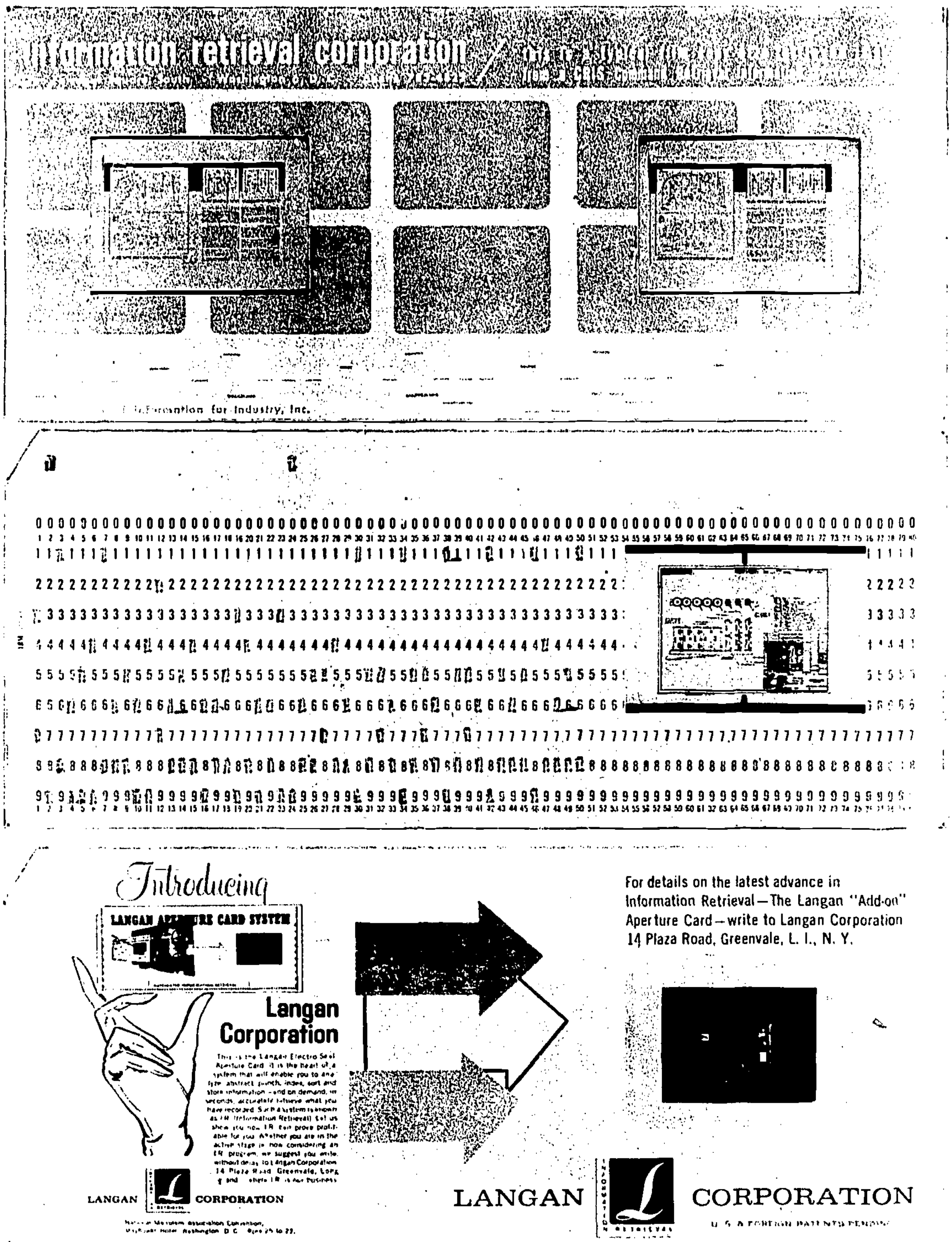


FIGURE 21. *Aperture microfiche. a. Information Retrieval Corp. CRIS Output. b. Aperture Card—IBM Corp. c. Aperture Card—Langan Corp.*

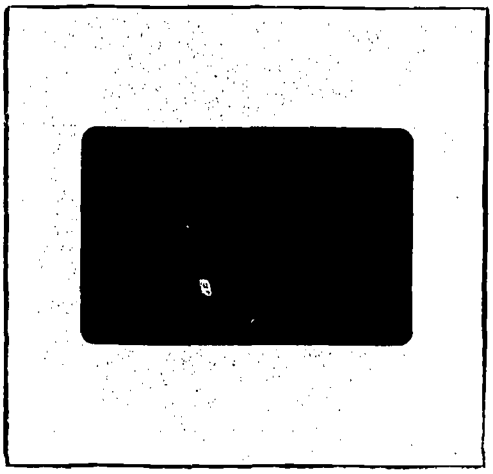


FIGURE 22. *Slide microfiche.*

images. However, polychromatic or full-color techniques are being explored. The chromatic capability of the storage is strongly dependent on the camera and film used for input recording. For most of the systems mentioned in this paper the lack of color is not important, because they are usually concerned with the storage of textual and black-and-white pictorial information. However, as color reproduction becomes more economical, it will be useful for pictorial prints, bar charts, and representations of color effects (spectograms, stress/strain patterns in materials, mixtures of crystalline materials viewed with polarized light, and the like).

In figure 15 and in the system descriptions in appendix A, the capacity of individual systems is indicated by the number of images per storage unit. This characteristic is best expressed, as is the system storage density, by the number of images that may be stored per cubic foot when loaded into their usual holder or magazine. This is a critical factor, especially for those systems in which the storage units have to be manipulated manually. For

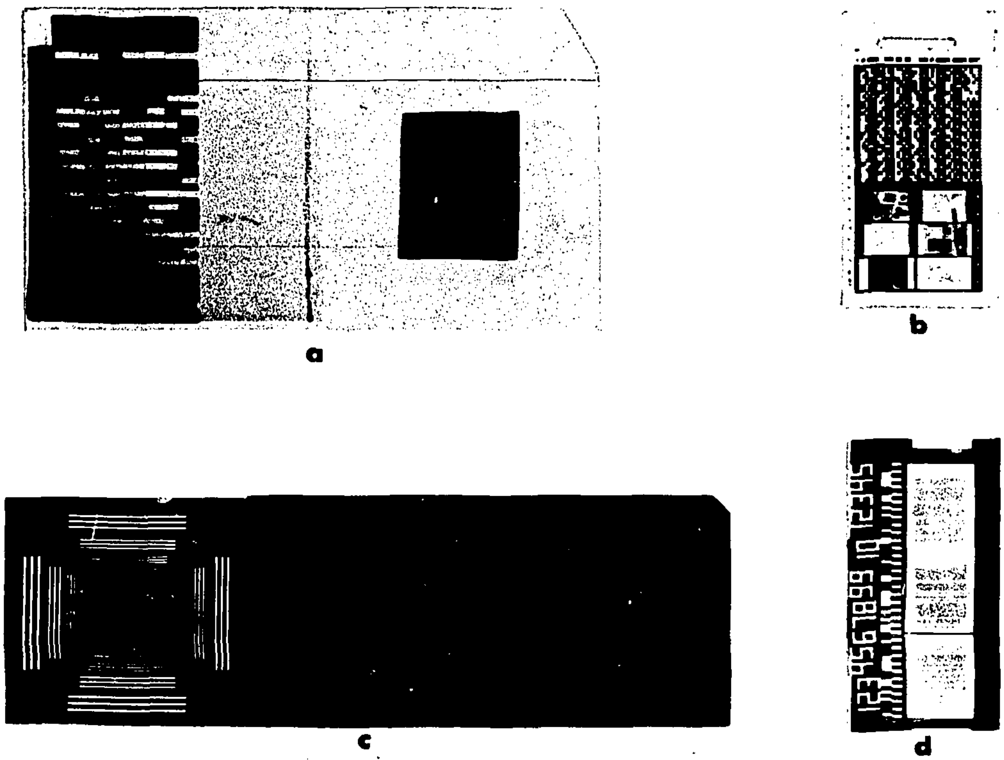
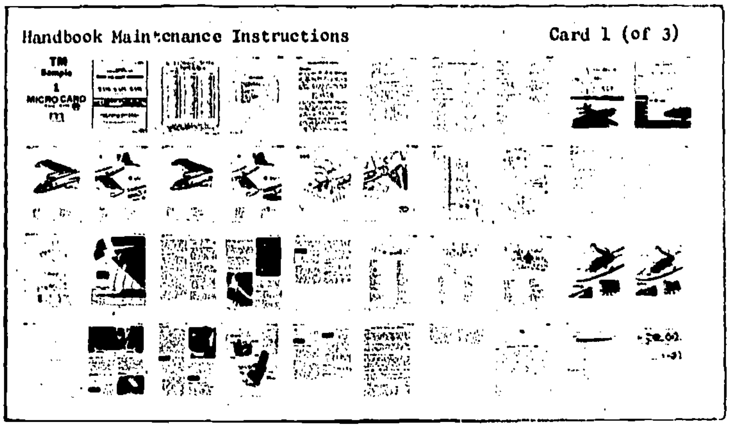
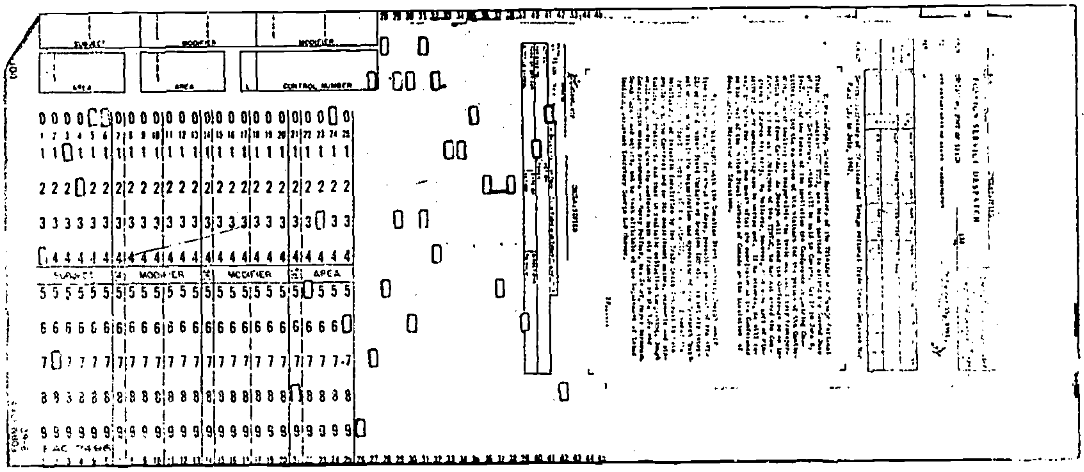


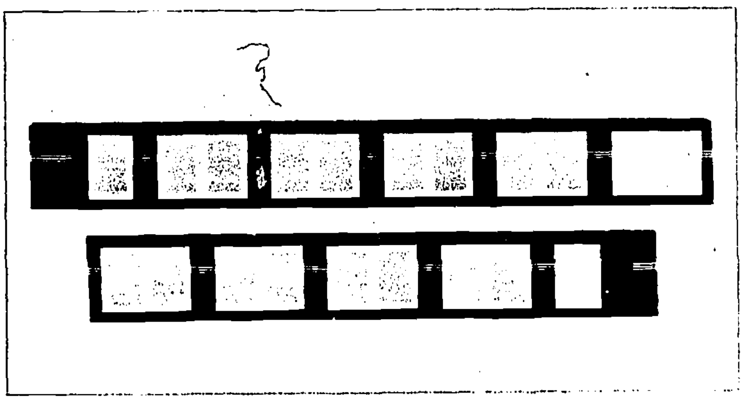
FIGURE 23. *Chip microfiche. a. Filmorex, Jacques Semain. b. Minicard, Eastman Kodak Co. c. Magnavue (includes magnetic coding area), Magnavox Co. d. Media, Magnavox Co.*



a



b



c

FIGURE 24. *Opaque microforms. a. Microcard—Microcard Corp. b. Electrostatic Print—Bell and Howell Corp. "DARE." c. Microtape (includes Kodamatic Indexing)—Recordak Corp.*

machine-manipulated units, this factor is significant primarily in terms of camera input requirements and in the mechanical design considerations. This image packing density, however, is deceptive for those systems that store more than one copy of the same image. In the Minicard system, for example, there may be 20 copies of the same document and this redundancy obviously must be considered in determining the effective capacity, i.e. unique documents stored, of the system.

The flexibility of a system is reported in terms of the ability (1) to update stored information by adding new information to individual documents, (2) to change information within a document by replacing old with new information, and (3) to add documents to or purge them from a storage file unit. Libraries may be most concerned with this matter of adaptability. However, some libraries may not wish to pay for the capabilities indicated above since their predominant activity may be to increase the total collection by adding new storage units and to purge from the collection by retiring those that have a low rate of usage because the information contained in them is outdated.

Replication media have been briefly discussed above. In review they may be classed as follows:

1. Transparency or translucency (plate or pliable film)
   a. Microfilm (pliable film)
      1. Roll (8, 16, 35, and 70 mm are the most common widths)
      2. Strip (an easily handled strip from a roll of film)
      3. Scroll (roll of film 10 to 20 inches wide)
   b. Microfiche (plate or pliable film)
      1. Unit (plate or pliable film usually of index card dimensions)
      2. Chip (usually pliable film containing one or a very few pages)
      3. Sheet (usually pliable film containing a number of documents)
      4. Jacket (usually pliable film put into acetate or card stock containers)
      5. Slide (plate or pliable film in metal, card stock, or plastic mounting)
      6. Aperture (usually pliable film mounted in card stock)
2. Opaque (card or paper stock)
   a. Microcard (emulsion on card stock)
   b. Microlex (emulsion on paper stock)
   c. Microprint (offset print on paper stock)
   d. Microtape (emulsion on heavy paper stock with pressure-sensitive backing)
   e. Electrostatic print (xerography on card or paper stock)
3. Electronic (magnetic pliable tape or card, or rigid disk)
   a. Video magnetic tape (pliable mylar tape 2 inches wide)
   b. Magnetic card (pliable mylar card—digital recording)
   c. Magnetic disk (rigid disk—digital recording)
   d. Magnetic tape (pliable mylar tape 1/4 to 1 inch in width—digital recording)

Replication methods show equally as much variety, as the following list indicates. (Original production methods, such as letterpress and typewriting, have not been included.)

1. *Electrostatic.* A dry copy process that requires no permanent negative and produces copy in a very short time (1/2 to 2 minutes). One-to-one full-size copying utilizes a low-resolution technique (about 10 lines per millimeter), but microcopy techniques are being developed that even now can give a 12 percent reduction.
2. *Facsimile.* A dry process using a charge potential that scans a piece of paper and, in its scan, arcs through and hence burns the paper. This has at present a very low resolution and is used primarily for the communication of full-size one-to-one copy by radio wireless or land-line communication systems.
3. *Microscan.* Not strictly a new reproduction process, this employs "a mask containing uniform arrangement of microscopic dots placed one-hundredth of an inch apart, allowing pages of information to be 'piled' on top of one another on a single sheet of film." (See item 61.)

4. *Offset printing.* Ordinarily a one-to-one printing process in which a master is prepared and copies of the original are made by having ink adhere to sensitive areas of the master, transferred to a blanket and then offset to paper. A microform modification of this technique is found in Microprint, in which 100 page images are printed onto a 6 by 9 inch sheet of paper.

5. *Photochromic.* A quite recent development by the National Cash Register Co. that can place over 2,500 page images on a unit microfiche of index card size by the radiation exposure of a thin organic film. (See item 20.)

6. *Photosensitive emulsions.* Except for full-size copy, picture snapshots, and the low-resolution one-to-one copy processes (e.g. blueprint, whiteprint, etc.), these are the most commonly used of the current microreplication means. It is normally a 2-step process that is dry or semidry for other than the silver halide emulsions, the latter being the most prevalent for microreproduction. Except as noted, resolution in these emulsions is quite good, with 200 lines per millimeter rather easy to obtain and up to 1,500 lines per millimeter available.

7. *Spirit and stencil duplication.* A dye transfer process involving the preparation of a master as with offset printing, normally used only for full-size reproduction as the resolution for microform purposes is quite poor.

8. *Thermography.* A single-step process in which a heat-sensitive emulsion is exposed to the carbon-base ink areas of a document by radiation from an infrared or heat source. Resolution for microform purposes is quite poor.

9. *Thermoplastic recording.* A recent development of the General Electric Co. The resolution for microform purposes appears to be good, about 130 lines per millimeter. The medium, a thermoplastic film, is heat-distorted in the process with the resulting recorded film being viewed through a special type of optical system. (See item 32).

10. *Video magnetic tape.* A magnetic recording on 2-inch mylar film generated by the electro-optical scanning of a desired image, this development in graphic recording techniques is only a few years old and may find application in the microform storage field. An electro-optical televisionlike system is necessary for display.

Except as noted, microform copies from these replication methods have to be viewed through various types of enlarging optical systems. Microreproduction techniques and viewing equipment are discussed much more fully in two rather comprehensive books by Ballou and Lewis (see items 3 and 52).[22]

Implications to the General Library

In the last section, it was noted that the more ambitious and forward-looking application of microform storage and the mechanized retrieval of graphic information have thus far been made outside of general libraries. (See item 57.) For some years, the Microcard Foundation and University Microfilms Inc., have made available rare and scholarly materials, theses, dissertations, unclassified AEC reports, and the like in microform. The provision of display and copy equipment in special libraries, information centers, and repositories gives access to microform copies of many of the unpublished U.S. Government and contractor reports. More and more, as agencies such as NASA begin to merge indexes and microcard copies of the informally reported material with the published literature in special fields of interest in their announcements, librarians in general libraries will also become interested. (See item 84.)

Well-known examples of microform storage of general interest are the *New York Times* on microfilm for use with a combination of a viewer/printer and the *Thomas' Register of American Manufacturers* on microfiche. Precursors of primary publication in microform are the experimental

[22] For further material on microreproduction see items 10, 43, 44, 51, 59, and 78.

productions of the journals, *Wildlife Disease*, and the recently announced *Statistical Methods in Linguistics*.[23] As more detailed information is disseminated on the operating experience with the recent generation of microstorage equipment, it may be feasible to apply the cost-effectiveness ratio for this equipment to conventional libraries. With such added insight, the necessarily budget-minded library administrators may attempt wider application of this technology.

Application Trends.[24]—In addition to cost experience, we must better understand the factors that govern patron acceptance of wider use of microforms. Naturally, normal habits are least affected whenever a full-scale reproduction is provided for the user. Even in this case, however, with presently available equipment, the legibility often leaves much to be desired. Moreover, this would hardly be a feasible approach to "purposeful browsing." However, it does not appear beyond the state of the art to produce acceptable copy from an original good microform. The problem is rather one which involves both economics and work discipline.

Assume, for the present, that in the foreseeable future economic considerations will make it unreasonable to respond to each inquiry with full-scale reproductions of the requested material. If we ignore the patron's natural reluctance to change well-ingrained work habits, the major issue of his acceptance of microform depends on achieving a reasonable balance between cost and quality of acceptable microform viewers. The illegibility and fatigue that underlie some of the adverse reaction of the patron are thus seen to be directly related. Besides legibility and freedom from abnormal fatigue, there remain some additional considerations that depend upon the form of the microfacsimile. It would be desirable to provide a function equivalent to that of quickly turning pages in a book, of quick reference back and forth among several documents, and eventually of annotating the item being studied.

In what follows, it is assumed that microform viewers can be improved to the point that would justify widespread patron acceptance. This is a reasonable premise since such improvements appear to be both technically and economically feasible. Under this assumption, one can speculate on the economics of responding to each patron request with an expendable microform, after the patron has properly identified the item he desires. Indeed, it looks as though the technology will eventually support a system in which this kind of service might be a less costly procedure to the library than the total costs now associated with storage of and accounting for loans of the original items.

The advent of inexpensive portable viewing equipment for personal use (an example of which is the Microcard Reader Mark IV), and the fact that some desirable material is available only in microform, points up the possibility that microform is beginning to compete seriously with conventional full-sized documents. The situation will be biased further when it becomes both economic and legal to supply the requester with a personal microform copy of the items he needs.

The role of the man with the crystal ball is not an enviable one in these days of rapid change. Yet the essence of planning involves some estimation of the effects of changes that are seriously in prospect. The continued growth in literature available through conventional publication techniques has greatly increased the base of potential holdings for a library. The difficulty of living with a much slower growth in working space and budget is posing a serious dilemma for library management. The obvious impracticality, even for major libraries, of simply increasing their holdings has led to several types of adjustment in response to the amazing increase in the number and variety of publications.

Several alternatives for coping with this problem are described below. These alternatives are listed roughly in the order in which they are considered to be acceptable to and within the reach of the average library management:

1. The library seeks to retain its subject matter coverage by a more selective sampling of the documents that are available.
2. The library deliberately restricts its subject matter coverage and becomes a more or less specialized collection in order to retain adequate coverage in depth for a selected portion of its patrons.

[23] This journal, published in Sweden, will be offered in English, French, and German, and in both full-size and microfiche editions.
[24] See references 1, 21, 25, 37, 41, 42, 55, 57, 67, and 75.

3. Groups of libraries agree to maintain coverage in depth in complementary subject areas and to depend on developing an effective interlibrary loan system for coverage of a wide range of subjects to cope with a wide range of reader interests.
4. The library acquires certain of its holdings in microform only and adapts to the curtailment of conventional services to certain of its patrons.
5. To supplement the interlibrary arrangement (see alternative 3), provisions are made for facsimile inspection, by means of direct communication, at one library of documents held in another library. To accommodate high priority need, there may be ancillary facilities for making copies immediately following such remote inspection.

Holding the less active portion of a library collection in microform is an emerging practice. Probably the same approach will be used for expanding the collection into new areas, particularly now that some serials can be obtained in microform. One may also expect a growing tendency for publications to be issued simultaneously in microform and hard copy. The decision regarding the form in which to acquire the holdings may be left to the library management. They might even consider holding journals unbound for a limited number of years and, after this period, retaining the journals only in microform rather than as bound volumes. The economics of this arrangement appears favorable.

Since the use of video transmission of graphic material is still in the early trial stage, it is not likely to be a widely considered alternative until the economics can be better estimated. Video does offer attractive features for situations in which it is not practical to make, in advance, facsimile copies of a large collection of infrequently used items. This browsing capability of video facsimile might provide the justification for experiments from which cost estimates can be derived for this kind of interlibrary loan.

In any event, recent trends show that the advantages of the first three alternatives are being actively exploited and that they appear to be reaching the point of diminishing returns for improving the effectiveness of a library's services to its patrons. As long as the number of publications continues to increase and as modern promotional techniques are used to bring new publications to the potential user's attention, there will be increasing pressure for the library to make more use of the fourth alternative—that of microform holdings. Libraries should, therefore, make plans for the eventuality that all major publishers will offer their products in both hard copy and microform.

Once technology makes it feasible for the library to provide the serious requester with a microform copy of selected portions or the entire contents of a document, adjustments in the use of copyright privileges, or even modifications of the copyright legislation, may be needed to make this alternative freely available to the library. There is a reasonable prospect that publishers of journals will find it advantageous to offer complete series in microform. If this microform came directly from the copyright holders some aspects of the touchy problem of copyright might be alleviated. Even so, realization of the full potential of microform is presently clouded by copyright problems. These problems are now being explored, and a serious effort to initiate action is in prospect.[25]

Needed Research and Testing.—As the improved microform art becomes more closely coupled to the technology of machine searching, the working procedures of both patron and librarian may have to undergo considerable adjustment. These technical changes will raise further questions of user needs and acceptance. Will the user adjust to different searching and read-out systems? How will the trade-offs in various systems

[25] While the copyright law of 1909 permits in spirit the "fair use" of copyrighted material, it is still a technical violation, however much ignored, to even handscribe (much less photocopy) copyrighted material (see items 15, 16, 28, 31, 50, and 85 in the bibliography). Specific recommendations for revision of the copyright law have been made by an ad hoc group called the Committee to Investigate Copyright Problems Affecting the Communication of Scientific and Educational Information. A study of the incidence of photocopying copyrighted materials was made by George Fry & Associates and reported in the publication *Survey of Copyrighted Material Reproduction Practices in Scientific and Technical Fields*, which was released for a very limited distribution in June 1963 by the National Science Foundation. The entire study was reprinted in the *Bulletin* of the Copyright Society of the United States of America for December 1963 (v. 11, no. 2: 69–124). John C. Koepke, one of the principal investigators, has reviewed the main findings of this study in a recent article, "Implications of the Copyright Law on the Dissemination of Scientific and Technical Information" (*Special libraries*, v. 54, Nov. 1963: 553–556).

affect his use, e.g. how far could one go in sacrificing quality of copy for speed before meeting user resistance? When the user discovers that a new system can provide him with information hitherto unavailable he may enlarge his demands for service. Can the system be designed to be flexible enough to take advantage of these changing user requirements? If we can determine user needs objectively, it would aid in efforts to improve technical facilities and, at the same time, allow cultivation of new working habits in the user.

Certainly, there is much that still needs to be determined about the relative virtues, to the user, of the address vs. the search types of microform retrieval systems. As mentioned previously, the search system requires a completely separate search procedure that should in turn, yield a convenient means for activating a mechanized microform retrieval device. The search may be accomplished by a separate logical manipulator that can be either a data processor of adequate capacity or a skilled reference librarian. Both will need access to appropriate indexes and other library tools.

The search systems are activated by inserting search identifiers directly into the microform retrieval device. Depending upon the nature and extent of these identifiers, this provides a capability for certain classes of search prescriptions. The extent to which it is advisable to incorporate computerlike logic to meet the user's need for selection from the microform file is not yet evident.

These have been presented as being the only alternatives for using a microform retrieval device. If automatic operation without human inspection during the selection processes is the desideratum, then they do represent alternatives. If the microform retrieval device has features for display of content to the user, then another mode of selection can be employed in which the user can participate in progressive modification of the selection criteria as well as control the selective copying operation that usually accompanies a display feature. Since it is technically feasible to combine the various features, it is likely that composite systems for microform retrieval will be used in the immediate future, since the selection process using a data processor has not been attempted in an operationally acceptable form.

The foregoing discussion points to the importance of evaluative pilot tests in carefully planned situations with appropriate analysis of the findings. The resulting information should then be made available in a useful form for guiding library management in the consideration of equipment. Related research and development topics that ought to be considered and included in such tests when technology has arrived at an appropriate stage of development include:

1. Investigation of fast but relatively inexpensive dry methods of reproduction in both microform and in full-size hard copy.
2. Development of relatively inexpensive means for introducing color in both the storage and reproduction of graphic material in microform.
3. Development of improved forms of facsimile communication and presentation, particularly with respect to the resolution and brightness of the display.
4. Investigations of the potential inherent in video tape, thermoplastic recording, and related means for storing a video facsimile.
5. Investigation of the potential in converting pictorial information into an entirely digital record so that it may be recovered for presentation or for producing photocopy, as well as utilized as the basis for machine inspection of the textual and pictorial information contained in the document.[26]
6. Evaluation by a combination of analysis and pilot testing to derive useful indications of the cost-effectiveness ratio of a system designed to provide access to archives of microform storage held in comprehensive centralized repositories.

Although the field is highly specialized, a wealth of experience may accrue from the present intensive research activities on command and control systems. The most important byproduct of these activities will probably be their impact in the area of human factors. There is particular

[26] For example, it is now possible to enter information in the format of orthographic projections into a computer and to process this information. One form of the processed result is a three-dimensional representation which can be displayed on a television-like image. This provides an elegant prospect for inspecting the contents of a computer file that has been derived from conventional engineering drawings.

emphasis on devising the man-machine interface in such a way that information can flow readily to and from the man. Here the critical requirement is immediate response, and this will probably be of interest in the future when mechanized versions of card catalogs, journal indexes, and similar library tools have been developed. Then a proper design of the man-machine interface will be needed in order to negotiate the inquiry in both form and content, with the collateral possibility of some measure of "browsability." While these are dramatic prospects, they probably lie well into the future.

On the other hand, there is the concurrent development of supporting systems to handle reconnaissance and intelligence data. These systems handle more data than do the command and control systems, but there is a slower pace at the man-machine interface. This pace, however, might be more appropriate to library usage and, in fact, the collection of and reference to information from overt sources have many aspects that parallel both general and special library needs. Much of the technology under discussion here has received support for research and development because of specialized needs to fit requirements of discrete collections of materials. Thus, these intelligence activities may have more immediate impact on library operations than the more dramatic investigations in command and control systems.

Economic Determinants.—Certainly cost will have much to do with the rate at which microform technology will extend further into library practice. Library management is caught in a sort of price squeeze of its own in adjusting to the growth of the literature and the closely related increase in demands for services. Heretofore, microstorage has been attractive primarily from the point of view of saving space without necessarily offering an economic advantage. A steady trend in the reduction of material costs and an increase in the productivity of microform equipments are beginning to change this situation.

The economic factors involved in conversion of holdings to microform have been the subject of study from time to time. A recent paper reports that a modest cooperative effort among a few libraries is sufficient to lower conversion costs to the point where they are offset by the value of the released storage space. (See items 24 and 64.) This savings is particularly significant when new construction for housing growing collections is under consideration.

If it is an acceptable practice to allocate the funds required for binding journals to pay for the purchase of them in microform, then it may not even be necessary to seek outside cooperation in order to have an economically viable situation. Libraries may also find it advisable to purchase back-issue journals in microform from such sources as University Microfilms and the Microphoto Division of Bell and Howell, rather than film their own copies.

Although microform is not new, it can now, for the first time, really be considered as an alternative to hard copy both from the viewpoint of benefits to the user and economic factors. This is so because only now has technology advanced so that not only are microform readers increasingly acceptable, but mechanized searching techniques coupled with high-quality photographic reduction permits the library to consider microform systems as a new approach to information control rather than just as a storage medium.

Even though cost considerations tend to determine immediate courses of action, the eventual acceptance of microform techniques will be determined much more by the cost-effectiveness ratio. Effectiveness is difficult to define and measure, since this would require an examination of the characteristics of library services and their contribution to the intellectual activities of our nation. These intellectual activities spread over a wide spectrum which extends from the cultural activities of the arts and literature at one end to the utilitarian aspect of science and technology at the other. There is some expectation that we may partially formulate criteria of effectiveness with respect to the utilitarian end of the spectrum. At present, there seems little prospect of finding a tractable approach with respect to the cultural end. The projected magnitude of public expenditures in support of scientific, technological, and related educational activities is large. This situation lends emphasis to the need to provide meaningful measures of effectiveness, if only to justify the increased library budgets required to support these activities.

Obviously, a measure of effectiveness is not an independent quantity that can be separated from the needs of the library user. His needs, in turn, derive in large part from the economic and sociological environment in which he works. Combining measures of effectiveness with the relatively more tractable cost determinations into a composite cost-effectiveness evaluation requires carefully planned and executed pilot tests with an objective examination of the results. In addition to thorough publication of such studies there ought to be carefully prepared demonstrations to convey the findings to those whose work environment may differ from the pilot test situation. By these means, the results of the cost effectiveness determination may be extended to other areas of interest. Thus, the testing and demonstration steps

Appendix A: FACSIMILE STORAGE AND

Name	Developer or manufacturer	Status [2]	Type	Size [3]	Purpose	Response time	Integration function	Input size	Storage media
Army Tactical Operations Central—ARTOC	Aeronutronics Division, Ford Motor Co.	O, N	Search	Large	Special—military field system.	Immediate	On-line.	Per camera	Slide microfiche.
Automatic Image Retrieval System—AIRS[9]	Recordak Co˜p.	O, N	Search	Medium	Special—INA	Delayed	Off-line	<11" x 34"	Roll microfilm.
Automatic Image Retriever.	Houston-Fearless Co.	O, C	Address	Medium	Special—INA	Delayed	Shunt	INA	Slide microfiche.
Automatic Minimatrex	Jonkers Business Machines, Inc.	D	NA	Medium	Special—index only.	Delayed	Shunt.	9" x 11" (digitally coded card only).	Strip microfilm.
Command Retrieval Information System—CRIS.[10]	Information Retrieval Corp.	O, C	Address	Medium	General	Delayed	Off-line	Per camera	Scroll microfilm.
Data Bank	Benson-Lehner Corp.	P	Search	Medium	General	Delayed	Off-line	INA	Aperture microfiche.
Dept. of Defense Damage Assessment Center—DODDAC.	Thompson Ramo Wooldridge, Inc.	O, N	Search	Large	Special—defense information system.	Immediate	On-line	Per camera	Chip microfiche.
Document Abstract Retrieval Equipment—DARE.	Micro-Data Div., Bell & Howell	O, N	Search	Medium	Special—abstracts only	Delayed	Off-line	<9" x 14"	Electrostatic print.
Documentary S'orage and Retrieval Sys'em.	Henry Staats	D	Search	Medium	General	Delayed	Off-line	<8½" x 11"	Unit microfiche.
Ececetron	Marcel Locquin	D	Search	INA	General	Delayed	Off-line	INA	Roll microfilm.
E-Z Sort with Aperture Insert.	E-Z Sort Systems, Ltd.	O, C	Search	Small	General	Delayed	Off-line	Per camera	Aperture microfiche.
Fast Access, Coded Small Images—FACSI.	FACSI, Inc.	O, C	Search	Small	Special—edge-notched card system of Society for Non-Destructive Te ting Journal	Delayed	Off-line	Page size	Microprint
FILESEARCH	FMA, Inc.	O, C	Search	Medium	General	Delayed	Off-line	<8½" x 14"	Roll microfilm.

See footnotes at end of table.

should properly be considered as an extension of the research program to provide some badly needed guidelines for the library manager.

Acknowledgments

The authors welcome this opportunity to express gratitude to their colleagues who participated in the preparation of this report. The primary source of material was NBS *Technical Note* 157 by Thomas C. Bagg and Mary Elizabeth Stevens cited as item 2 in the bibliography. Their pertinent ideas helped to bring out the issues to be considered in achieving wider acceptance of this particular technology in the operation of general libraries.

RETRIEVAL SYSTEM DESCRIPTIONS [1]

Storage coding	Storage unit capacity	Storage density (in images per cu. ft.)	Selection	Access time [4]	Output [5]	Printout time [6]	System flexibility			
							Update	Change	Add	Purge
Magnetomechanical integral index.	1,000/system.	INA [7]	Automatic, magazine	1.5 sec.	Display	NA [8]	Yes.	Yes.	Yes.	Yes.
Photoelectric integral index.	2,500/reel	2.4×10^5	Automatic, magazine	7.0 sec.	Display; hard copy	25.0 sec.	No.	No.	No.	No.
INA	8,000,000/ system.	INA	Automatic, magazine	0.3 sec.	Display; hard copy	INA	Yes.	No.	Yes.	Yes.
Optical integral index.	1,000,000/ strip.	1.0×10^6	Semiautomatic	5.0 min.	Display; punched paper tape, magnetic tape and/or hard copy.	INA	Yes.	No.	Yes.	Yes.
Electro-optical integral index.	500,000/ scroll.	INA	Automatic, magazine	20.0 sec.	Display; film aperture card.	15.0 sec.	No.	No.	No.	No.
Magneto-optic integral index.	75,000/ rack	3.5×10^4	Semiautomatic, magazine.	2.0 min.	Display; hard copy	INA	No.	No.	Yes.	Yes.
INA	200/magazine.	INA	Automatic, magazine	30.0 sec.	Display; copy	Included in access time.	Yes.	Yes.	Yes.	Yes.
Electromechanical integral index.	NA	INA	Semiautomatic	INA	Duplicate microimage card.	INA	No.	No.	Yes.	Yes.
Electrical integral index.	180/plate	5.4 x 10[illegible]	Automatic magazine	12.0 sec.	Display; microimage copy.	INA	No.	No.	Yes.	Yes.
Electro-optical integral index.	INA	INA	Automatic	1.0 min.	Display; copy	INA	No.	No.	No.	No.
Mechanical integral index.	INA	INA	Semiautomatic	INA	Per viewing equipment.	NA	No.	No.	Yes.	Yes.
Mechanical integral index.	8/card	INA	Semiautomatic	INA	Display	NA	No.	No.	Yes.	Yes.
Photoelectric integral index.	32,000/reel	2.9×10^5	Automatic	2.5 min.	Display; 35mm roll microfilm, 3M hard copy.[11]	6.0 sec., 20.0 sec.	No.	No.	No.	No.

See footnotes at end of table.

Appendix A: FACSIMILE STORAGE AND

Name	Developer or manufacturer	Status [1]	Type	Size [2]	Purpose	Response time	Integration function	Input size	Storage media
Film Library Instantaneous Presentation—FLIP.	Benson-Lehner Corp.	O, C	Address	Medium	General	Delayed	Off-line	INA	Roll microfilm.
Film Optical Scanning Device for Input to Computers—FOSDIC II.	National Bureau of Standards.	O, N	Search	Medium	Special—data systems only.	Delayed	Shunt	EAM card only.	Roll microfilm.
Film Optical Scanning Device for Input to Computers—FOSDIC IV.	National Bureau of Standards.	O, N	Search	Medium	Special—data systems only.	Delayed	Shunt	EAM card only.	Roll microfilm.
FILMOREX	Jacques Semain	O, C	Search	Small	General	Delayed	Off-line	<8½″ x 11″	Chip microfiche.
FILMSORT (with EAM equipment).	Remington Rand.	O, C	Address	Small	General	Delayed	Off-line	Per camera	Aperture microfiche.
Graphic File and Retrieval System.	Itek Corp.	O, C	Search	Large	Special—engineering drawings.	Delayed	Shunt	Per camera	Chip microfiche.
Hi-Speed Color Printer.	Radio Corp. of America.	D	Search	Large	Special—intelligence system.	Immediate	On-line	Per camera	Slide microfiche.
Intellofax	Central Intelligence Agency.	O, N	Address	Small	General	Delayed	Off-line	Per camera	Aperture microfiche.
Keysort with Microform Inserts.	Royal-McBee Corp.	O, C	Search	Small	General	Delayed	Off-line	Per camera	Jacket or aperture microfiche.
LODESTAR with Image Control Keyboard.	Recordak Corp.	O, C	Address	Small	General	Delayed	Off-line	Per camera	Roll microfilm.
LODESTAR with Kodamatic Indexing.	Recordak Corp.	O, C	Address	Small	General	Delayed	Off-line	Per camera	Roll microfilm.
MAGNAVUE	Magnavox Co.	D	Search	Large	General	Delayed	Off-line	<28″ x 28″	Aperture microfiche.
MEDIA	Magnavox Co.	O, C	Address	Medium	General	Delayed	Off-line	<8½″ x 14″	Chip microfiche.
METRICARD Analysis Console with Computer.	Thompson Ramo Wooldridge, Inc.	INA	Search	Large	Special—photo intelligence processing.	Immediate	On-line	Per camera	Chip microfiche.
MICROCARD System	Microcard Corp.	C, C	Search	Small	General	Delayed	Off-line	Per camera	Microcard
MICROCITE II	National Bureau of Standards.	O, N	Search	Medium	Special—abstracts only.	Delayed	Off-line	<3″ x 5″	Sheet microfiche.
MICROCITE II, Model 2	National Bureau of Standards.	D	Search	Medium	Special—abstracts only.	Delayed	Off-line	<3″ x 5″	Sheet microfiche.
Microfilm Finder—Reader System.	Massachusetts Institute of Technology	D	Address	Small	General	Delayed	Off-line	INA	Roll microfilm.
Microfilm Storage and Retrieval System.	General Precision Laboratories. Mosler Safe Co.	O, C	Search	Medium	General	Immediate (with) shunt. Delayed (with) off-line.		Per camera	Aperture microfiche.
Micro Image Locator	National Bureau of Standards.	O	Address	Medium	Special—limited information.	Delayed	On-line	Generally small.	Sheet microfiche.
MICROLEX File	Lawyers Cooperative Publishing Co.	O, C	Address	Small	Special—law books only, at present.	Delayed	Off-line	Book pages	Microlex

See footnotes at end of table.

RETRIEVAL SYSTEM DESCRIPTIONS—Continued

Storage coding	Storage unit capacity	Storage density (in images per cu. ft.)	Selection	Access time [4]	Output [5]	Printout time [6]	System flexibility			
							Up-date	Change	Add	Purge
Photoelectric integral index.	7,200/reel	INA	Automatic	2.0 min	Display	NA	No	No	No	No.
Photoelectric on document image.	12,000/reel	1.3×10^6	Automatic	1.2 min	Duplicate of original EAM card.	0.1 sec	No	No	No	No.
Photoelectric on document image.	12,000/reel	1.3×10^6	Automatic	1.2 min	Duplicate of original EAM card, magnetic tape.	0.1 sec., INA	No	No	No	No.
Photoelectric integral index.	4,000/drawer	1.5×10^4	Automatic, magazine	1.0 min	Display; original chip.	Auxiliary	No	No	Yes	Yes.
Electromechanical integral index.	Varies	INA	Automatic	INA	Display; aperture card itself.	Auxiliary	Yes	No	Yes	Yes.
Magnetic separate and photoelectric integral index.	INA	INA	Automatic	INA	Display; duplicate chip.	INA	Yes	Yes	Yes	Yes.
Electromechanical integral index.	80/magazine	INA	Automatic	1.0 min	Map with current information.	1.0 min	Yes	Yes	Yes	Yes.
Visual integral index	INA	INA	Manual	INA	As desired with auxiliary equipment.		No	No	Yes	Yes.
Mechanical integral index.	Various	INA	Semiautomatic	INA	As desired with auxiliary equipment.		No	No	Yes	Yes.
Photoelectric counting index.	2,500/reel	2.4×10^5	Automatic, magazine	5.0 sec	Display; hard copy	25.0 sec	No	No	No	No.
Visual counting index.	2,500/reel	2.4×10^5	Semiautomatic, magazine.	10.0 sec	Display; hard copy	25.0 sec	No	No	No	No.
Magnetic integral index.	30,000/block	2.3×10^4	Automatic, magazine	3.8 min	Display; system aperture card, hard copy.	INA	Limited.	Limited.	Yes	Yes.
Photoelectric integral index.	200/capsule	INA	Semiautomatic	1.0 min	Display; hard copy	INA	No	No	Yes	Yes.
Visual and photoelectric integral index.	60/magazine	INA	Automatic or manual, magazine.	INA	Display	NA	Yes	Yes	Yes	Yes.
Visual integral index	80/card	2.3×10^5	Manual	INA	Display; microcard itself.	NA	No	No	Yes	Yes.
Optical separate index integrated for search.	18,000/system.	7.5×10^4	Semiautomatic	15.0 sec	Display; filmstrip, snapshot (Polaroid).	1.0 sec., 10.0 sec.	No	No	No	No.
Optical separate index integrated for search.	18,000/sheet.	7.5×10^4	Semiautomatic, magazine.	15.0 sec	Display; filmstrip, snapshot.	1.0 sec., 10.0 sec.	No	No	No	No.
Visual stroboscopic integral index.	INA	INA	Semiautomatic	INA	INA	INA	No	No	No	No
INA	5,000/drum	INA	Automatic, magazine	4.0 sec	Display	NA	No	No	Yes	Yes.
Electromechanical synchronization.	10,000/sheet.	INA	Automatic	2.0 sec	Copy, photopaper	Included in access time.	No	No	No	No.
Visual integral index	INA	INA	Manual	INA	Display	NA	No	No	Yes	Yes.

See footnotes at end of table.

735-898 O—64——10

Appendix A: FACSIMILE STORAGE AND

Name	Developer or manufacturer	Status [2]	Type	Size [3]	Purpose	Response time	Integration function	Input size	Storage media
Micro Research System.	Petroleum Research Corp.	O, C	Search	Small Medium	Special—as now available. General.	Delayed	Off-line	<8½″ x 11″ and strip charts.	Unit microfiche.
MINICARD	Eastman Kodak Co.	O, C	Search	Large	General	Delayed	Shunt	<8½″ x 14″	Chip microfiche.
MINIMATRIX	Jonkers Business Machines, Inc.	O, C	NA	Medium	Special—index only.	Delayed	Off-line	9″ x 11″ only	Strip microfilm.
MIRACODE	Recordak Corp.	O, C	Search	Medium	General	Delayed	Off-line	Per camera	Roll microfilm.
Photochromic Micro-Image System.	National Cash Register Co.	O, C	Address	INA	General	Delayed	Off-line	Per camera	Unit microfiche.
Photo-Magnetic System.	Peter James	P	Search	INA	General	Delayed	Off-line	<8½″ x 14″	Roll microfilm, magnetic tape.
Random Access Document Indexing and Retrieval—RADIR.	Hallicrafters Co.	O, C	Search	Medium	General	Delayed	Off-line	<34″ x 44″	Roll microfilm.
RAP 600	System Development Corp.	O, C	Address	Small	Special—teaching machines and the like.	Delayed	Off-line	INA	Slide microfiche.
Rapid Access Look-Up System.	Ferranti-Packard Electric, Ltd.	O, C	Address	Medium	General	Delayed	Off-line	INA	Roll microfilm.
Rapid Selector	National Bureau of Standards.	O, N	Search	Medium	General	Delayed	Off-line	<22″ x 34″	Roll microfilm.
Seventy Millimeter Selector.	Photo Devices, Inc.	D	Search	Medium	General	Delayed	Off-line	<31″ x 31″	Roll microfilm.
Unitized Microfilm System.	Xerox Corp.	O, C	Address	Medium	General	Delayed	Off-line	<40″ x 50″	Aperture microfiche.
VERAC	AVCO Corp.	O, C	Address	Medium	General	Delayed	Off-line	INA	Sheet microfiche.
Video File System	Radio Corp. of America.	D	Search	INA	General	Delayed	Off-line	<8½″ x 11″	Video tape (electronic).
WALNUT	International Business Machines Corp.	O, N	Search	Large	General	Delayed	Shunt	<8½″ x 14″	Strip microfilm, also magnetic index (electronic).

[1] This is a comprehensive list of all systems which could be identified; therefore the following systems, for which no descriptive information was made available, deserve mention: Litton (developed by Litton Industries, Inc.), Microprint File (developed by Readex Microprint Corp.), Rapid Random Access Protector, Recall Film Index System, Spectral Data Card, Target Map Coordinate Location, and Viewer Reproducer.

[2] Letters in this column are defined as follows: O—operational, C—commercial, N—noncommercial, D—developmental, P—proposal.

[3] Systems are designated as follows: Small—less than $10,000; medium—$10,000 to $200,000; large—more than $200,000.

[4] Access time refers to the time required to get a display image.

[5] Display refers to an impermanent reproduction on a reader or console screen; copy is used to designate some kind of microreproduction or hard-copy output; specific information about the type of copy, e.g. punched paper tape, is given when available.

RETRIEVAL SYSTEM DESCRIPTIONS [1]—Continued

Storage coding	Storage unit capacity	Storage density (in images per cu. ft.)	Selection	Access time [4]	Output [8]	Printout time [6]	System flexibility			
							Up-date	Change	Add	Purge
Mechanical integral index.	140,000/ drawer.	1.2×10^5	Semiautomatic	3.0 min	Display; film card, hard copy.	INA, 15.0 sec.	No	No	Yes	Yes.
Photoelectric integral index.	24,000/ stick.	6.8×10^5	Automatic, magazine	1.0 min	Display; duplicate film chip.	2.0 sec	No	No	Yes	Yes.
Optical integral index.	100,000/ strip.	1.0×10^6	Manual	10.0 min	Display	NA	No	No	Yes	Yes.
Electronic functionally integral index.	2,500/reel	2.4×10^6	Automatic, magazine	INA	Display; hard copy	25.0 sec	No	No	No	No.
INA	2,625/plate	INA	Semiautomatic	INA	INA	INA	No	No	No	No.
Magnetic separate index.	INA	INA	Automatic	INA	Hard copy	INA	Yes.	Yes	No	No.
Photoelectric from separate index.	10,000/reel	INA	Semiautomatic, magazine.	2.0 min	Display: negative Kalvar filmstrip or aperture card.	2.0 min	No	No	No	No.
INA	000/ systems.	INA	Semiautomatic, magazine.	7.0 sec	Display	NA	No	No	Yes	Yes.
Photoelectric integral index.	880/reel	INA	Automatic, magazine	3.0 sec	Display	NA	No	No	No	No.
Photoelectric integral index.	36,000/reel	9.6×10^4	Automatic	6.0 min	35 mm roll microfilm	On the fly	No	No	No	No.
Photoelectric integral index.	3,200/reel	1.5×10^5	Automatic	5.0 min	Display, 3M hard copy.	20.0 sec	No	No	No	No.
Visual integrel index.	INA	INA	Manual	INA	Display; 24" x 36" Xerox.	INA	No	No	Yes	Yes.
Mechanical integral index.	1,000,000/ systems.	INA	Automatic, magazine	2.0 sec	Display; microfilm	0.5 sec	No	No	No	No
Magnetic integral index.	36,000/reel	1.4×10^5	Automatic	5.0 min	Display; electrofax hard copy.	7.0 sec	Yes.	Yes	Yes	Yes.
Magnetic separate index.	990,000/ module.	1.4×10^4	Automatic, magazine	12.0 sec	Hard copy, aperture card.	20.0 sec., 10.0 sec.	No	No	Limited.	Limited.

[6] Printout time refers just to the time required to produce the hard copy: the total lapsed time from access to hard copy would be the sum of access time and printout time. When two printout times are listed they refer respectively to the different output forms available.

[7] INA—information not available.

[8] NA—not applicable.

[9] This system has been succeeded by MIRACODE described below.

[10] This system is the successor to AMFIS.

[11] This component is manufactured by the Minnesota Mining and Manufacturing Co.

APPENDIX B

Bibliography

Note: References to the many systems described in Appendix A have not been included in this bibliography. Those readers who would like to become better acquainted with them are referred to the second item which contains a comprehensive bibliography covering such systems through 1961.

1. Artandi, Susan. Special library services—current thinking and future trends. Special libraries, v. 54, Feb. 1963: 103–106.
2. Bagg, Thomas C., *and* Mary E. Stevens. Information selection systems retrieving replica copies; a state-of-the-art report. [Washington, U.S. Dept. of Commerce, National Bureau of Standards] 1961. 172 p. ([U.S.] National Bureau of Standards. Technical note 157)
3. Ballou, Hubbard W., *ed.* Guide to microreproduction equipment. 2d ed. Annapolis, Md., National Microfilm Association, 1962. 519 p.
4. Battelle Memorial Institute, *Columbus, Ohio. Dept. of Economics and Information Research.* Specialized science information services in the United States; a directory of selected specialized information services in the physical and biological sciences. Washington, National Science Foundation, Office of Science Information Service, 1961. 528 p. (NSF 61–68)
5. Bedford, Gwendolyn M. Review of current machine systems for handling information. Paris, North Atlantic Treaty Organization, 1956. 17 p. (North Atlantic Treaty Organization. Advisory Group for Aeronatical Research and Development. Report 46)
   ASTIA document no. AD–138 063.
6. Bello, F. How to cope with information. Fortune, v. 62, Sept. 1960: 162–167, 180, 182, 187, 189, 192.
7. Bibliography in an age of science. Urbana, University of Illinois Press, 1951. 90 p. (Phineas L. Windsor lectures in librarianship, 1950)
8. Bishop, Charles. Problems in the production and utilization of microfiche. American documentation, v. 12, Jan. 1961: 53–55.
9. Blumberg, Donald F. Information systems and the planning process. New York, Diebold Group [n. d.]
   Paper presented at the First Joint Meeting of the Canadian Operations Research Society and the Institute of Management Science, Toronto, Ontario, Canada.
10. Born, Lester K. The literature of microreproduction, 1950–1955. American documentation, v. 7, July 1956: 167–187.
11. Bourne, Charles P. Bibliography on the mechanization of information retrieval. Menlo Park, Calif., Stanford Research Institute, 1958. 22 p.
   ——— ——— Supplement. 1+ Menlo Park, Calif., Stanford Research Institute, 1959+
12. Bourne, Charles P. The historical development and present state-of-the-art of mechanized information retrieval systems. American documentaion, v. 12, Apr. 1961: 108–110.
13. Bowker, K., *and others.* Technical investigation of elements of a mechanized library system. Cincinnati, AVCO Corp., 1960. 110 p. (AVCO Corporation. Final report no. RW–6680)
14. Brannen, George B. A literature survey of technical information services. Special libraries, v. 54, Feb. 1963: 94–101.
15. Brown, Alberta L. Summary of copyright positions. Special libraries, v. 52, Nov. 1961: 499–505.

16. Budington, William S. Using copyrighted material. Special libraries, v. 52, Nov. 1961: 510–513.
17. Burkig, J., *and* L. E. Justice. Magnacard—magnetic recording studies. *In* Western Electronic Show and Convention. WESCON convention record. [Papers] v. [1] pt. 4; 1957. [New York] Institute of Radio Engineers. p. 214–217.
18. Bush, Vannevar. As we may think. Atlantic monthly, v. 176, July 1945: 101–108.
Condensed and illustrated version appeared in *Life*, v. 19, Sept. 10, 1945: 112–114, 116, 118, 121, 123–124.
19. Bushor, William E. Information storage/retrieval. Electronics, v. 35, June 29, 1962: 40–62.
20. Carlson, C. O., D. A. Grafton, *and* A. S. Tauber. The photochromic micro-image memory. *In* Symposium on Large-Capacity Memory Techniques for Computing Systems, *Washington, D.C., 1961*. Large-capacity memory techniques for computing systems; [proceedings] Edited by Marshall C. Yovits. New York, Macmillan [1962] (ACM monograph series) p. 385–412.
21. Chase, James W. Library 21: American Library Association exhibit at Seattle World's Fair. Special libraries, v. 53, July/Aug. 1962: 339.
22. Chronic, John. How microfilm library aids research. World oil, v. 142, May 1956: 95–97.
23. Clapp, Verner W. Library photocopying and copyright: recent developments. Law library journal, v. 55, Feb. 1962: 10–15.
24. Clapp, Verner W., *and* Robert T. Jordan. Reevaluation of microfilm as a method of book storage. College and research libraries, v. 24, Jan. 1963: 5–15.
25. Current research and development in scientific documentation. no. 1+ July 1957+ Washington, National Science Foundation, Office of Scientific Information. semiannual.
Semiannual reports containing descriptive statements from individuals and organizations undertaking research projects in scientific documentation.
26. Davis, Watson. Microphotographic duplication in the service of science. Science, v. 83, May 1, 1936: 402–404.
27. Devlin, Thomas J., *and* W. T. King. Technical correspondence: control and retrieval through microfilm and punch card techniques. Special libraries, v. 51, Oct. 1960: 421–424.
28. Dorward, Donald. A publisher looks at copyright. Special libraries, v. 52, Nov. 1961: 505–510.
29. Doss, Milburn Price, *ed.* Information processing equipment. New York, Reinhold Pub. Corp., 1955. 270 p.
30. Ellsworth, R. S. New horizons with microfilm. American documentation, v. 2, Oct. 1951: 221–228.
31. Freehafer, Edward G., *and others.* Joint Libraries Committee on Fair Use in Photocopying: report on single copies. Special libraries, v. 52, May/June 1961: 251–255.
32. Glenn, W. E. Thermoplastic recording. Journal of applied physics, v. 30, Dec. 1959: 1870–1873.
Also issued as General Electric Research Laboratory reprint 3350.
33. Greenough, M. L. New uses of microfilm with electronic scanners, a progress report on FOSDIC III. *In* National Microfilm Association. Proceedings of the 8th annual meeting. 1959. Annapolis, Md. p. 279–286.
34. Grolier, Eric de. International Advisory Committee for Documentation and Terminology in Pure and Applied Science; a preliminary report. [n. p.] 1955. 31 p.
Unpublished report in UNESCO working papers series, no. 320/5601.
35. Hayes, R. M. The magnacard system. *In* International Conference for Standards on a Common Language for Machine Searching and Translation, *Western Reserve University, 1959*. Information retrieval and machine translation; based on the International Conference for Standards on a Common Language for Machine Searching and Translation. Editor: Allen Kent. New York, Inter-Science Publishers, 1960. (Advances in documentation and library science, v. 3) v. 1, p. 563–574.

36. Hayes, R. M., *and* J. Wiener. Magnacard—a new concept in data handling. *In* Western Electronic Show and Convention. WESCON convention record. [Papers] v. [1] pt. 4; 1957. [New York] Institute of Radio Engineers. p. 205-209.
37. Heiliger, Edward. Application of advanced data processing techniques to university library procedures. Special libraries, v. 53, Oct. 1962 : 472–475.
38. Heilprin, L. B. Communication engineering approach to microforms. American documentation, v. 12, July 1961 : 213–218.
—— —— *In* National Microfilm Association. Proceedings of the 10th annual meeting. 1961. Annapolis, Md. p. 80–92.
39. Heller, Elmer W., *and* Charles D. Hobbs. A survey of information retrieval equipment. Santa Monica, Calif., System Development Corp., 1961. 30 p. (SP–642)
40. Information retrieval. Chemical and engineering news, v. 39, July 17, 1961 : 102–110, 112; July 24 : 90–96, 98.
41. Johnson, H. Thayne. An approach to the library of the future. Special libraries, v. 53, Feb. 1962 : 79–85.
42. Karth, Joseph E. Libraries in the space age. Special libraries, v. 53, Oct. 1962 : 462–465.
43. Kiersky, Loretta J. Bibliography on reproduction of documentary information. Special libraries, v. 51, Feb. 1960 : 72–77; v. 52, Mar. 1961 : 132–136; v. 53, Mar. 1962 : 135–140.
44. Kiersky, Loretta J. Developments in photoreproduction. Special libraries, v. 51, July/Aug., Dec. 1960 : 306–307, 554–555; v. 52, Apr., July/Aug., Dec. 1961 : 188–189, 320–321, 581–582; v. 53, July/Aug., Dec. 1962 : 331–332, 608–609.
45. King, G. W., G. W. Brown, *and* L. N. Ridenour. Photographic techniques for information storage. *In* Institute of Radio Engineers. Proceedings of the IRE, v. 41, Oct. 1953 : 1421–1428.
46. Kuipers, J. W. Microcards and microfilm for a central reference file. Industrial and engineering chemistry, v. 42, Aug. 1950 : 1463–1467.
47. Kuipers, J. W. A research program on information searching systems. *In* ADIA—Conference, *Frankfurt am Main, 1959.* Proceedings. Frankfurt/Main, Deutsche Gesellschaft für Dokumentation, 1961. (Beiheft zu den Nachtrichten für Dokumentation, Nr. 8) p. 58–69.
48. LaHood, Charles G., *Jr.* Production and uses of microfilm in the Library of Congress Photoduplicatoin Service. Special libraries, v. 51, Feb. 1960 : 68–71.
49. Langan, John F., by mesne assignment to Film 'N File, Inc., New York. Record Card. U.S. Patent 2,512,106. Patented June 20, 1950; filed January 3, 1946. Class 40–158
50. Latman, Alan. Copyright Office recommendations for a new copyright law. Special libraries, v. 52, Nov. 1961 : 514–521.
51. Lewis, Chester M. The interrelationship of microfilm, copying devices and information retrieval. Special libraries, v. 53, Mar. 1962 : 130–134.
52. Lewis, Chester M., *and* William H. Offenhauser, *Jr.* Micro-recording—industrial and library applications. New York, Interscience Publishers, 1956. 456 p.
53. Lowry, W. Kenneth. Some functions, interactions and problems of communication. Special libraries, v. 51, Nov. 1960 : 479–482.
54. Luther, Frederic. Microfilm, a history 1839–1900. Barre, Mass., Barre Pub. Co., 1959. 195 p.
55. Marr, Donald. Planning for library and company future needs. Special libraries, v. 51, Apr. 1960 : 188–190.
56. McKenna, F. E. Readin', ritin', and reproducin'; tools for the special librarian. Special libraries, v. 53, Nov. 1962 : 526–530.
57. Mooers, Calvin N. The next twenty years in information retrieval: some goals and predictions. *In* Joint Computer Conference. Proceedings. 1959. New York, Institute of Radio Engineers. p. 81–86.
——— ——— Cambridge, Mass., Zator Co., 1959. 18 p. (Report no. ZTB-121; AFOSR TN 59–245)
ASTIA document no. AD 212 225.
58. Mueller, Max W. An evaluation of information retrieval systems. Burbank, Calif. Lockheed Aircraft Corp., 1959. 114 p. (Lockheed Aircraft Corp., Burbank, Calif. Memorandum report no. MR-7170)

59. National Microfilm Association. Proceedings of the 7th–10th annual meeting. 1958–1961. Annapolis, Md. 4 v.

60. Nelson, A. M., H. M. Stern, *and* L. R. Wilson. Magnacard — mechanical handling techniques. *In* Western Electronic Show and Convention. WESCON convention record. [Papers] v. [1] pt. 4; 1957. [New York] Institute of Radio Engineers. p. 210–213.

61. New reproduction technique. Graphic science, v. 4, Mar. 1962 : 25.

62. Parker, F. M. Engineering drawing processing system. Special libraries, v. 51, Oct. 1960 : 429–432.

63. Power, Eugene. Microfilm as a library tool. Special libraries, v. 51, Feb. 1960 : 62–64.

64. Pritsker, Alan B., *and* J. William Sadler. An evaluation of microfilm as a method of book storage. College and research libraries, v. 18, July 1957 : 290–296.

65. Rabinow, Jacob. Presently available tools for information retrieval. Electrical engineering, v. 77, June 1958 : 494–498.

First presented at the American Institute of Electrical Engineers Summer General Meeting, Montreal, Canada, June 24–28, 1957, as conference paper 57-860.

66. Richmond, Phyllis A. What are we looking for?—attention to the nature of scientific discovery would produce better information retrieval systems. Science, v. 139, Feb. 22, 1963 : 737–739.

67. Schultheiss, Louis A., Don S. Culbertson, *and* Edward M. Heiliger. Advanced data processing in the university library. New York, Scarecrow Press, 1962. 388 p.

68. Seidell, Atherton. The photomicrographic reproduction of documents. Science, v. 80, Aug. 24, 1934 : 184–185.

69. Sharp, Harold S. Information retrieval—pitfalls of information retrieval. Industrial research, v. 3, Apr./May 1961 : 33.

70. Shaw, Ralph R. Machines and the bibliographical problems of the Twentieth Century. *In* Bibliography in an age of science. Urbana, University of Illinois Press, 1951. p. 37–71.

Reprinted under title *Bibliographic Organization* by the University of Chicago Press (1951).

71. Shaw, Ralph R. Mechanical storage, handling, retrieval, and supply of information. Paris, North Atlantic Treaty Organization, 1956. 34 p. (North Atlantic Treaty Organization. Advisory Group for Aeronautical Research and Development. Report, 50)

——— ——— Libri, v. 8, no. 1, 1958 : 1–48.

ASTIA document no. AD 144 216.

72. Sophar, Gerald J. Micro-opaques. Special libraries, v. 51, Feb. 1960 : 59–62.

73. Staats, Henry M. Data extraction in nondestructive testing. Nondestructive testing, v. 15, Jan./Feb. 1957 : 44–46.

74. Sullivan, Walter. Electronic microfilm archive is seen as an aid for overcrowded libraries. New York times, Aug. 9, 1959, p. E9.

75. Swanson, Don R. Library goals and the role of automation. Special libraries, v. 53, Oct. 1962 : 466–471.

76. Symposium on Information Storage and Retrieval Theory, Systems, and Devices, *Washington, D.C., 1958.* Information storage and retrieval theory, systems, and devices. Edited by Mortimer Taube and Harold Wooster. New York, Columbia University Press, 1958. 228 p. (Columbia University studies in library service, no. 10)

77. Tate, Vernon D. An ingenious searcher. PMI, photo methods for industry, v. 4, July 1961 : 66–67.

78. Tate, Vernon D. Microreproduction. PMI, photo methods for industry, v. 4, Sept. 1961 : 35, 99, 104.

79. Tauber, Maurice F. Problems in the use of microfilm, microprint, and microcards in research libraries. Industrial and engineering chemistry, v. 42, Aug. 1950 : 1467–1468.

80. Tyler, A. W., W. L. Myers, *and* J. W. Kuipers. The application of the Kodak Minicard system to problems of documentation. American documentation, v. 6, Jan. 1955 : 18–30.

81. U. S. *Congress. Senate. Committee on Government Operations.* Documentation, indexing, and retrieval of scientific information; a study of Federal and non-Federal science information processing and retrieval programs. Washington, U.S. Govt. Print. Off., 1960. 283 p. (86th Cong., 2d sess. Senate. Document no. 113)

82. U.S. *Dept. of Commerce. Advisory Committee on Application of Machines to Patent Office Operations.* Report. Washington, Dept. of Commerce, Sales and Distribution Division, 1954. 76 p.
Vannevar Bush, chairman.

83. U.S. *National Science Foundation. Office of Science Information Service.* Nonconventional technical information systems in current use. no. 1+ Jan 1958+. Washington, U.S. Govt. Print. Off.

84. U.S. *President's Science Advisory Committee.* Science, government, and information: the responsibilities of the technical community and the Government in the transfer of information; a report. Washington, U.S. Govt. Print. Off., 1963. 52 p.

85. Varmer, Borge. The copyright law revision. Special libraries, v. 52, Apr. 1961: 185–188.

86. Walker, Fred L., *Jr.* Blueprint for knowledge. Scientific monthly, v. 72, Feb. 1951: 90–101.

87. Warheit, I. A. Machines and systems for the modern library. Special libraries, v. 48, Oct. 1957: 357–363.

88. Warheit, I. A. The microfiche. Special libraries, v. 51, Feb. 1960: 65–67.

89. Worsley, Peter K. Data retrieval with especial application to use of Film Library Instantaneous Presentation (FLIP) in literature searching. *In* Los Angeles. University of Southern California. *School of Library Science.* Modern trends in documentation; proceedings of a symposium held at the University of Southern California, April 1958. Edited by Martha Boaz. London, New York, Pergamon Press, 1959. p. 70–73.

90. Worsley, Peter K., *and others.* A study of the fundamentals of information storage and retrieval. Los Angeles, Benson-Lehner Corp., 1959. 97 p. (Benson-Lehner Corp., Los Angeles, Calif. Final report no. 417; Contract Nonr-2666(00))
ASTIA document no. AD 229 709.

91. Young, George V., Inglewood, Calif., assignor of one-half to Peter J. Pohl, Manhattan Beach, Calif. Video image frame recording and reproducing system. U.S. Patent 2,955,157. Patented Oct. 4, 1960; filed Aug. 22, 1956. Class 178–6.6.

CONFERENCE SESSION IV

Libraries and Automation

RUTHERFORD D. ROGERS
Library of Congress

Before we start this morning's discussions, I would like to do a little stocktaking. I want to try to express the point of view of librarians for the benefit of the computer people who are here. At the same time, I am going to try to assess what has been said so far in the conference for the benefit of those who are not technical experts. I will address myself first to problems which are bothering librarians and for which we are seeking help from the computer technology.

Let us begin with the processing end of library science. Certainly, most research libraries are worried about arrearages. If they do not actually have arrearages—and I think most libraries do—they are concerned about keeping up with the control of their collections. It would be wonderful if computers could speed processing or somehow simplify it so that it took less manpower. In this same vein, we are concerned about the wasteful duplication of effort among research libraries, about the fact that so many of us are doing the same job. One of the reasons for this is that processing is slow and expensive with the result that present efforts are not prompt or comprehensive enough to satisfy everyone.

The largest research libraries have problems keeping track of serials; such libraries are unable to keep up with claims for missing issues, and they are not as sure as they should be that they are getting what they are paying for. Even in monographic literature we librarians have occasional difficulty making sure that we do or do not have something. This is partly a result of not having all our records up to date. Therefore we run the risk of ordering items that we already have.

We are concerned about the depth of indexing. I know that there will not be universal agreement on this even among people at the Library of Congress. Perhaps we would be reasonably satisfied with our control of monographic materials if processing were up to date, but we certainly do not have the control in depth over serial literature that we would like to have.

We are concerned about the size and the complexity of the card catalog. It is expensive to maintain; the bigger it gets, the more expensive it gets. Size slows down filing and makes it hard to find things. We are concerned about the complexity of our notation system for classification and relative shelving. Too many mistakes are made in putting long and involved call numbers on the spines of books. This same complex notation is difficult to manage in shelving books with the result that a lot of items get misplaced so that for all intents and purposes books are lost.

On the subject of relative shelving, I suggest that a good many people who call themselves scholars are deluded into thinking they can really do research work by browsing. There is no doubt that there are certain things one can accomplish by browsing. But anyone who is half a librarian—or half a scholar—knows that there is no one place or no two or three places in a big collection where one is able to get everything needed on a subject. From that standpoint, the bibliographical approach is much sounder than the browsing approach.

We are also concerned about maintenance of subject heading lists and classification schedules. If hearings are clearly out of date or if we do not have an appropriate heading for a new subject, then to that extent our tools are weakened and the tools of all libraries that rely on our system are weakened. With present methods, the size of the staff and the publishing costs needed to keep subject heading lists revised and reissued as frequently as they should be are fairly monumental.

We are concerned about space, which is the bête noire of all large research libraries. Perhaps there is a solution for this in microreproduction, but I believe that the space problem is related to some other problems that are perhaps much deeper than just square footage. We should be concerned about eliminating the redundancy in our collections, and also the unused materials. I know these are both very dangerous statements because we all know from experience that there are books which lie unused in research libraries for years and then become vitally necessary. But if you approach this problem from the standpoint of redundancy you will agree, I believe, that we have a terrible layering of the same information and that frequently one book would be just as good as another for the research worker's purposes. Furthermore, the redundancy in our collections complicates the administration of research libraries in organizing and keeping track of these materials.

Perhaps even more serious, particularly if we were to become automated, is the possibility of inundating the individual reader or user of the library with a superfluity of material. This is a subject about which we are already hearing a great deal.

Finally, we are concerned with the speed of response, not only the speed of response for readers but also for internal processing—searching as part of the acquisition and ordering functions, the establishment of entries, and similar activities.

Now, if I have understood what has happened so far, the specialists have told us that we are not yet at the point where we can feasibly store the intellectual contents of all the books and documents of a large research library in a computer. We are going to discuss graphic storage this morning, but this is just a variation of what we are already doing. We can store the bibliographical approach to the collections and, as I understand it, this means that we can not only put the National Union Catalog into the computer but also other catalogs. In so doing we have a record of the Library of Congress and of each contributing library in the computer store.

I do not believe we said very much about another possibility, that of having the computer store bibliographic information now issued in book form. Yesterday Dr. Taube said that we may be approaching the time quite soon when some of the big abstracting services will not publish in book form. The only way that we will then be able to benefit from these services is to obtain what they do in some machine-readable form and get it into our computer system; by having it in our system we would be able to speed the access to the entire record.

It has not been claimed that we can speed or even facilitate the actual indexing of serial material by machine, although I would hope that a computer might make this possible some day. (I am thinking now of the work that Swanson and others have done in machine indexing.) The computer does, however, hold out a definite promise for managing the tremendous bibliographical apparatus that would be required if we are to have control of individual serial articles in essentially the same place and in the same manner that we control monographs. The absence of this control, I think, is one of the big weaknesses of our present system; certainly scientists are increasingly dissatisfied with it for this reason. I believe it has been promised that automation can speed up our access to the store, although there seems to be a difference of opinion as to whether or not we are going to get instantaneous response, or whether there will have to be a delay of some duration in order to batch requests.

I would hope that a computer might make it possible to simplify establishment of entries, even though librarians would still have to do a good deal of the descriptive cataloging. It does occur to me that one does not need quite as rigid a system as we now have for descriptive cataloging, simply because there could be so many different access points to a given document by virtue of the flexibility inherent in computer manipulation of data.

It is reasonable to expect that automation would improve our acquisition procedures in at least two fundamental ways: (1) by assuring that materials in the library are reflected promptly in the catalogs and (2) by making it possible to determine more rapidly whether we have a given item. A study that we made at the Library of Congress indicated that a searcher spends most of his time walking from one tray to another; it is not the time he spends after he pulls the tray out, but the transit time, that consumes his working day.

Stated another way, I would hope that computers would assist the research worker and the librarian to exercise effective command over the contents of a very large library. I doubt that any of us would claim that we are doing so at present. Furthermore, computers and modern communication channels and devices would make it possible for research workers at remote points to have access to a central store. This would mean, first of all, that it would not be quite as important that each library maintain separate card catalogs. If one had adequate access to a central catalog that had the records of the local library as well as those of other libraries and, secondly, if one could get from this central record, by reasonably inexpensive printout, a book-form catalog, the necessity of maintaining a multiplicity of local card catalogs would diminish. However, where such local catalogs were maintained, certainly the processing function could be completed much more rapidly by querying the central store.

Finally, I hope that, in effect, the computer experts are telling us that if we adopt computer technology, even though it is not perfect at this point, we will be setting the stage for much more important developments in the future, when we may actually put the intellectual content into the machine in digital form and manipulate it, when we can eliminate redundancy and furnish the reader with the information that he needs, perhaps not in book form but in capsule form, giving him what he is really seeking, regardless of the form in which it was originally published.

Review of Microforms: Preliminary Remarks

JOSEPH BECKER

Introduction

We are talking this morning about the paper which Sam Alexander and his group at the National Bureau of Standards prepared and which is designed to survey the current status of graphic storage techniques. I think it does this well. It describes in some detail the materials and the forms that exist for storage; it touches on the viewers to a certain extent; it dwells rather heavily on the systems and the equipment used for graphic storage—here its information is derived from some 50 or 60 questionnaires which were sent out to customers and manufacturers who either use or produce graphic storage equipment. The paper concludes with an indication of the research and testing needed in this field, and it also touches on several library problems. Now the field itself is rather technical, and I have chosen to review some of the technical terms with you so that in the discussion period we will be talking from the same foundation.

Microform Materials

Let's talk first about the material. When we speak of film today we no longer speak of just silver film; rather we are talking about a much larger family of materials. It is useful to know a little bit about these materials so that we have an appreciation of their capabilities.

Silver halide is quite common to us; we use it in our box cameras, and it is the first kind of film that was employed for graphic storage. Diazo followed. This is a dye material which is coated on a film or mylar base. By playing ultraviolet light over it you disintegrate certain portions of the diazo compounds and, where this disintegration does not take place, when you subject the material after exposure to ammonia vapor, it brings out the dye that remains in the coating. This imbeds itself into the film. So whereas silver has a layer which you can scratch, diazo imbeds itself into its basic layer and is not as susceptible to scratching as is silver. Diazo became quite attractive to film people because it was a relatively dry process; it needs only to go through this gaseous ammonia for development and does not require the wet chemicals and fixing that is customary with the silver.

Kalvar was a development which came a little later. It is the same mylar but coated with a collection of little gas bubbles which, again, is subjected to heat. The heat, which need only be that of a warm iron, will actually cause some of those little bubbles to break and form light scattering centers which result in the image which we see on the film. Kalvar is even more attractive from a developing viewpoint because it just requires heat; it is a dry form of copying. Kalvar film actually works through two little rollers that have a little heat coil in each one and they produce the image rather dramatically and very quickly.

Photochromics is an even newer technique which the National Cash Register Co. has been working on for the last several years. We are most accustomed to photochromics, in a sense, in the "no-carbon-required" paper. Forms, which used to have the carbon interleaved, no longer have the carbon because the verso of each form is coated with a chemical material, which consists of a collection of microscopic bubbles, and by pressing hard on the surface on the face of the first form you are breaking some of the bubbles. These bubbles contain dye which, when exposed to air, results in the image. The NCR people have actually coated film with the same type of substance. In this case, however, there is no breaking of these little bubbles, instead, by playing ultraviolet light on the film the bubbles change from a colorless state to a colored state, and this gives the resultant image. This technique has even greater power because you can erase the colored image with white light if you choose. So here then is a technique for recording information on film and then at a later stage removing an image or a line of an image as well. The developing processes here are dry so this makes it even more attractive.

Thermoplastic, a General Electric development, is newer than all the other techniques. It is a surface coated with a plastic material which is subjected to an electron beam that optically records graphic information on plastic film essentially by melting the plastic.

The last technique is video which is equivalent to recording information on video tapes, just as we do for our TV commercials and broadcasts. There has been some work done in this field, particularly and notably by RCA, but there are not very many systems, if any, that employ this technique at the present time.

Microforms

So, very quickly, that covers the materials in the field and it embraces what I consider the interesting one. The report then describes the microforms themselves, that is, the way in which we use these basic properties that I have described. The basic materials are used in one of two ways, and I think this report very interestingly classifies these two ways and helps us to remember the categories logically. These two basic classifications are the transparent or translucent group and the opaque group. In the translucent group are the conventional rolls of microfilm. In library work we started with 35 mm microfilm for journals, newspapers, and the like; 16 mm microfilm was introduced a little bit later. There are strip forms in which this translucent form can exist: this is nothing more than chopping up pieces of a reel of film and handling them that way. There is the scroll which is essentially a wider reel, but there are scrolls of individual images. These are again translucent or transparent and, in order to view them, you normally use projection techniques: you put the light through the image.

For storage purposes there are jackets which are little sleeves of transparent material so that as you photograph the basic data you can cut it up and slip it into a glassine sleeve and the full information is contained, for example, in a 5 by 8 inch area. This has the added advantage of permitting the addition of material to any given file if the need arises.

There is the sheet film which requires a step-and-repeat camera. The end result is a translucent 5 by 8 or 3 by 5 inch form (you pick the size), which is actually a sheet of film on which data have been deposited, image by image, in horizontal rows.

There is the window, or aperture, card which is made by taking one or two images at a time from the reel and mounting them, using some pressurized device, into a little adhesive-lined window; then you can reproduce as many as you want. There are some systems which employ up to eight 16 mm images on an IBM card so that you have the advantage of punched information on the left side of the card and can still store up to eight images on the right-hand side of the card. The Filmsort Co., a subsidiary of Minnesota Mining and Manufacturing Co., is probably the chief manufacturer of aperture cards.

And, finally, there are chips which are essentially slides or Minicards, which I'll come to later. The opaque form, the most common to the library usage, is the Microcard. In this medium the original data are recorded on film, and then the resultant 16 mm negative is contact printed to the reverse side of a 3 by 5 card. You can get from 30 to 60 or more images on the verso of a catalog card and use the front part of the card for recording bibliographic information. Microtape is similar but has the advantage of being on adhesive-backed material, and instead of 3 by 5 cards it is in strips. This gives you the opportunity then of snipping off your contact-printed material and adhering it to any kind of a document for storing this type of data.

Microprint is something which the Readex Co. in New York features. They produce a sheet of film which then burns an offset master from which one can actually produce paper microreproductions. This is another opaque form of storage. Finally, there is Microlex, which is an opaque sheet of film for storing records on both sides.

Very briefly, I have reviewed the two forms in which this material is kept for storage purposes: on a transparent or translucent medium, or on an opaque medium. In each case these microforms require viewers, and this has been a thorny problem for librarians because viewers (a) are expensive and (b) present some technical difficulties, not the least of which are the hot spots on the film and the high lights in the center of the projection screen as opposed to the requirement for high intensity of light in order to get a clear reflected image on a ground glass screen from an opaque form of storage.

Address-Type Microform Systems

The report then classifies the systems that have resulted over the years for storing some of these materials. Again, there are two basic classifications: "address-type" systems and "search-type" systems. The address-type system refers to the form of storage that requires only a number, or one identifier, in order to locate the material. The search-type system provides a facility for a greater number of selection criteria. Given a file of material in numerical sequence, with the address-type system one could retrieve data immediately if the item number is known. Search systems have digital recording associated with the document image so that one can do some Boolean operations, logical operations, and find data that way.

CRIS is a mechanized scroll that consists of thousands of images laid down side by side. Upon receiving an address, the machine locates the x, y coordinates of a given image and projects it on a screen. Media utilized the same approach except that the document is on a little chip about half the size of a 5-cent stamp.

Walnut is an IBM machine, a very expensive one, which uses Kalvar strip film on which are placed pairs of images. These strips are loaded into plastic cells so that there may be something like 10 or 20 strips to a cell and as many as 100 or more cells to a bin and the system can grow that way. In this system you can retrieve any image in any cell on any strip by knowing its address. When the address is specified, the machine cycles to the right plastic cell location, grabs the particular strip, pulls it up to the proper height, and shines ultraviolet light through the Kalvar strip onto an unexposed Kalvar aperture card which becomes exposed at that particular point. (Each Kalvar aperture card can hold about 8 images.) The card then moves through warm rollers that produce the final product ready for viewing. The Walnut machine can be addressed manually by hitting a series of numbers. It also can be hooked up to a computer which performs the intellectual operations and massages the basic data and comes up with a series of numbers which lead to the Walnut cell, or it will produce a collection of IBM cards which can then operate the Walnut machine.

There are examples of equipment associated with each of these systems. I have just briefly enumerated some of them; the report gives a good summary of all of them, although it doesn't describe them in much detail. FLIP, Film Library Instantaneous Presentation, built by the Benson-Lehner Corp. for the Air Force, is one of the first machines of its kind—a sequence finder for microfilm. It consists of a 2,000-foot spool of film containing something like 72,000 frames; when you ask for document number 65,321, for example, the machine automatically cycles to that particular location, and you then view the document on a screen.

Lodestar is a more recent machine produced by the Recordak Corp. of Eastman Kodak. This company has introduced one of the few novelties in this area in the last few years, and a rather successful one, namely, cartridge film. In the library world we have primarily been accustomed to using spools of film; these require threading, and we have concern about the film being scratched or otherwise damaged. Cartridge film, such as is used in home movie cameras, requires a very simple form of loading. Lodestar is the same type of thing. Where there is inherent indexing, as in our journals and newspapers, we can find something by knowing the journal title, date, and the page of a given article. If material is microfilmed in sequence, one then can zero in on a given page without much trouble. Lodestar recently hooked up a little device which looks like an adding machine keyboard; data can be retrieved when one types the six or seven numbers representing an image location on the 100-foot spool. You go in with a given address, punch the keyboard, and the machine locates the information for you at once.

AVCO Corp. has worked on an experimental device called VFRAC which reduces documents 100 to 1 (you are looking at the round head of a pin at that reduction) and deposits the images on a 10 by 10 inch glass plate. About 10,000 such images can be stored on the face of a plate; the idea is to get a well of these plates. Given the address of any one image, the machine would cycle to a particular glass plate, pull it up, and then a television scanner would come in and with three orthogonal motions locate the x, y, and z location of the particular image and display it on a monitor at some distant location or produce a hard copy. These then are the address systems.

Search-Type Microform Systems

The search-type systems are a little more complicated. We know them best in the library world by the one which Vannevar Bush and Ralph Shaw worked on in the middle forties. They wanted to put on 2,000-foot spools 35 mm film abstracts from the *Bibliography of Agriculture* with some digital code alongside. The object was to search by specifying code selections, after which the machine would then find the particular abstract, shoot right through it onto some unexposed film and, finally, provide the user with a small strip of film which contained the abstracts most pertinent to his particular request. Now they had technical difficulties in those days, mainly with acceleration and deceleration of that film; they just hadn't achieved it at the time. It used to spill out over the floor, particularly when there were several successive hits all in one location. So that was put on ice for a while, although in recent years the National Bureau of Standards reworked it and a descendant of the Rapid Selector is functioning now in the Bureau of Ships at the Navy Department in Washington.

Filesearch is probably the latter day sophisticated Rapid Selector. Here the manufacturer, FMA, Inc., has overcome all of the technical difficulties that plagued the Rapid Selector. This machine can perform many types of basic logical operations; it can find things on a reel of film quite well; it can produce images on a ground glass screen for viewing; and it can provide a print, a hard-copy enlargement, directly from the same device.

Minicard is probably the most sophisticated chip-type system ever produced. A whole family of equipment was designed to manipulate chips of information. Data is reduced at a ratio of 60 to 1, so that about half the area of a 5-cent stamp holds 12 pages of documents and a code equivalent to a full IBM card. This system requires special camera equipment to record the codes and the images at these reduction ratios simultaneously. It requires a chopper to put these things into their exact dimensions, a waxer that coats them in order to preserve them better, a sorter, a selector, and quite a bit of storage equipment. Now this is a rather elaborate system but it taught the profession a great deal about the associated technology.

The whole point was that whereas with linear systems—for example, Filesearch and Rapid Selector—you normally would have to go through the entire reel to locate a particular document, with the chip system you have a more or less random approach. You can locate individual items more readily if you have organized the store to begin with. Filmorex, invented by the Frenchman Semain, is the same type of idea (except that the chips are bigger).

Video File was introduced by RCA about a year ago. They have prototype equipment in Camden but it has not received too much attention within or outside the company for some time. The idea here was to scan documents with a video scanner, record them on 2-inch-wide magnetic tape spools, locate records digitally, and then exit them with video technique by getting images on a TV tube. On a magnetic tape one has the option of recording video information as well as digital information. The usefulness, of course, with Video and VERAC is that data can be communicated over great distances to remote locations.

Areas for Research

The report indicates that we need more research in the library world to understand these address systems and search systems and to recognize where each can be used profitably. We are not using search systems in research library environments today, that I know of; we are using address systems. We do this because our material is fairly well organized to begin with. Where heterogeneous material is to be organized, then search systems become candidates for consideration.

The report mentions man-machine interface. Here is the console again coming in, because if you have a search system that will permit logical combinations to be asked for then you have the same situation as you do when the individual wants to communicate with the closed system of the computer.

There seems to be an indication that we need better methods of reproduction and enlargement from film. Librarians are very conscious of this and they would like to get dry copies and enlargements that will be acceptable to their users. Color is another important area that deserves research because we have been concentrating mainly on black and white, and there are problems associated with recording and enlarging in color.

Remote communications is another area that has not been fully explored. Finally, there is needed research in the area of these new materials—thermoplastics and photochromics. What can these do for us now as storage media? What advantages can we derive from them that we cannot get from some of the more traditional media?

The report concludes by identifying some library problems, particularly as they relate to the continuing growth rate in libraries. The aim of microfilm has been to provide more compact storage of printed materials; with the rate of growth that we are experiencing now, space continues to be a problem. User acceptance, the question of legibility again, and user fatigue in prolonged use of microfilm on a viewing screen continue to be problems, as do the tough copyright difficulties that exist. Our copyright laws provide some rather serious constraints to ready, open, and easy copying. The question of durability and preservation needs to be considered. We have lived with film now for 40 or 50 years, but we still are not sure that, as a basic recording medium, it is permanent enough to satisfy the librarian.

We have had very little experience with cost data. This came up yesterday in regard to our regular library operations and the same thing is true here. We have had two good studies, one in 1957 by Pritsker and Sadler (reference 64, p. 139) and, more recently, one by Forbes and Waite for the Council on Library Resources, Inc., which was reported in *College and Research Libraries* by Verner Clapp and Bob Jordan (reference 24, p. 137). The conclusions of the latter articles were that with groups of libraries working together in the initial processing of material, cost advantages can be achieved that might not otherwise be available if libraries operated such activities on their own. And last, the question: Do we need more mechanization in this field or can we be content with what we have? I will leave this list of library problems on the board, throw the discussion open to the floor, and let you shoot at these.

General Discussion

WOOSTER: The report might have mentioned the possibility of Government standards for microfilm. I get the impression that AEC and NASA, within the last week or so, have issued a rather detailed specification indicating that from now on the Government will use an 18 to 1 reduction. Now, I couldn't care less whether it is 15.7, 16, or 18, as long as we finally get together. Those of you who have had the problem of buying viewers know that the things are essentially fixed field and do not permit that much adjustment.

Although it is impractical to talk about costs, one of the things which we all need answered is the question of whether to buy a camera or to hire service done on microfilming. We need illumination on that.

VOSPER: I would like to bring out a couple of fundamental misconceptions in the paper under discussion. One of these appears in the introduction [page 111], where the paper implies, although it doesn't make a great deal of it, an attitude of resistance on the part of librarians to all of these techniques we are discussing. I think this is a straw man. We can assume that with respect to computers and facsimile microsystems there is a sense of urgency and need on the part of librarians, and that both parties here today are trying very hard to move in the right directions. Actually, although two paragraphs later an apparently emotional or unreasoning concern about esthetics is discussed, I don't think it affects the total library attitude toward these techniques. However, one must accept the fact that at a certain point esthetics *is* a real problem. The report mentions the attractiveness of finely detailed maps. If one is talking to a geologist, this is a real problem, not an emotional problem. We are generally more concerned, I think, with certain imprecisions and inadequacies in both microtype and computers. These imprecisions and inadequacies the librarian must face realistically. If one talks about storing masses of serial literature, one must think of serving not only the geologist concerned with color but also a number of other needs that are imprecisely met with present systems.

In this paper the authors said that thus far the library community has not followed the lead of the business world in rushing into the use of new equipment. Here again I think there is a fundamental misconception about the economy we are discussing; after all the library community is not a business community. Many of the systems discussed in the paper are being developed as large governmental or commercial enterprises where risk capital and large sums of money are available. The economy of the average research library is of a completely different order. Even in my fairly well-to-do library, risk capital just does not exist, and one cannot undertake a system without pretty full assurance that it will work, that it will endure, and that it will be cheaper than the existing system. Furthermore, very few of us in the library world have access to research and development money. In my library we are trying to get some, but it is not a simple matter. This raises a fundamental difference that needs to be taken into account by both parties in the discussion. We are working in the same direction if we recognize these basic differences.

WAITE: With regard to cost studies, I was glad that you did call attention to Verner Clapp's article. This is, I believe, the starting point for anyone who wants to consider justifying microfilm in a library application for cost-vs.-storage reasons. Now of course the dynamics of the information problem are being recognized, so that it's doubtful that we have to justify everything on storage, but this opens Pandora's box in establishing values on all of these other benefits; we haven't been able to do this.

We ought to be cautious in applying microreproduction for the storage of graphic materials, as far as coupling microreproduction with automation and mechanization, in the library situation or, as a matter of fact, in almost any other situation which we have seen. Perhaps the easiest way to make this point is simply to say that we have found, in one of our recent studies, in a situation where the volume is perhaps the highest I have ever seen, that the only way to do the operation is manually because there isn't an automatic way that is fast enough.

CLAPP: As I see it, the great advantage of microcopy in library work is as a medium of publication. This is not to denigrate its great importance as a preservative, as a copying medium, as a method for avoiding costs of binding, as an intermediate between the original and, let us say, a Copyflo, and various other things of this kind. But its great potential and still unexploited characteristic is as a medium of publication. We have some examples of this, but not much more than examples. The International Geophysical Year was able to publish the meteorological reports in millions of microcards which otherwise would have taken millions of feet of shelves for publication in ordinary form. (Actually this made the difference between publishing and not publishing.) Micro has made it possible for any library to have English publications before 1640, American publications before 1800, and other rariora which *no* library, no matter how wealthy, can possess in the original, and of which no complete sets of the originals exist.

Here then is an enormous potentiality. We are hardly using it. The reason we are hardly using it is the high cost of micro! We are now paying, on a per page basis, as much for micro, and sometimes more, as we do for inkprint material. As I see it, in order to improve this situation we must go to higher ratio reductions. If we get 10,000 images in the place of one original, we've got a lower per page cost which makes publication and dissemination extremely attractive. "Okay," you say, "why don't we do it?" The answer is very simple—we can't read the stuff after it has been so distributed. There it stands 10,000 pages on one page in some device—a Walnut, or a Minicard—and there it is locked up in the machine, far away from the user. We need some quick, reliable, convenient, and inexpensive intermediary between the microstore and the reader beyond what we have now. Present reading devices are not satisfactory. In business and industry, which is favorably compared here to library work, they can afford to have well-devised, convenient, optically excellent machines which cost anywhere from $3,000 to $10,000, because a person is employed at a good annual wage to sit and read bank checks, engineering drawings, what have you. Nobody is going to give a 4- or 5-thousand-dollar-a-year clerk a machine into which he is going to have to squint, and dodge hot spots, and work in a dark room in a dusty corner of an antique library. Here there is commercial advantage; you can count the cost.

We can't do that in libraries, fiscally speaking, and I doubt if we could do so operationally speaking. I doubt if it would be useful to put an $8,000 reading device for microcopy into a library because in the first place our readers aren't trained to use it, and in the second place this is much better than they need. All they need is to find a certain page in the *New York Times* or a certain page in some 16th-century English publication. So we give them what they need, we give them a Model E or a Model C, or something like that, and let them dodge the hot spots. This is inconvenient; it is the best we can do; it just barely serves; it is limited. It will never promote the use of micro; it is better than nothing, and this is about all that can be said for it. To be really advantageous the system ought to permit this potential user to take away a copy. This is really the difficulty. The reader-printers will supply a copy, but it will be a copy at such a price that you have now *lost* the economic advantage of the reduced cost of dissemination. There is no point in disseminating material at one-tenth of a cent a page if, in order to read a page, you have to pay 10, 15, 25 cents for it, or even 6. The economic advantage is lost at 6 cents a page; try it on a 300-page book anytime you wish, and see how many times you want to give a reader a 300-page book blown up from microfilm. Not very often, and the reader won't very often be willing to pay for this himself.

What is the answer? I think the answer is a personal reading machine with which you can read micro as conveniently as you do the original. Now this was done in the 12th century by spectacles. In the 12th century, at the age of 40 people stopped reading. Then along came some monk or another and he made this contraption and now 60-year-old men go on reading and writing and *talking* about it. It does seem as though in the 20th century we ought to be able to do almost the same thing for microform. Somewhere, around one of these corners here, lies this little reading device which I can pull out of my hip pocket to read micro, I don't know at what ratio, whether 15.5, or 20, or 60, or maybe a 100, but the librarian will be able to hand out little strips at a fraction of a cent apiece and not at 10 or 15 cents per page for hard copy. At that time the use of micro in libraries will be liberated.

Now, let me just finally point out, having made

735-898 O—64——11

this lengthy address, the potentiality of this in solving some of the problems that Rudy Rogers has talked about. First, the reduction of size. If our major research libraries could agree on a stock of books to be reduced, to be eliminated, or to be sent to second-rate storage (in Herman Fussler's terms) maybe we can agree to reduce this by high-ratio reduction microphotography to a few file cases—retaining the catalog of it, retaining the ability to have access to it, actually making it more available than it was before because now every subscribing library will have copies instead of just having the material scattered around.

You see in the Human Relations Area Files, for example, the potentiality of organizing material by selection and microcopying. The same potentiality exists in a great many subjects. Albert Boni some years ago offered the American Chemical Society a service by which he would reproduce in microprint all the articles digested or abstracted by *Chemical Abstracts*, and he had the thing fairly well laid out. ACS bowed out because of the copyright problem. Copyright problems can always be licked. There is just a little matter of payment involved. Now we are back to potentialities again.

HEILPRIN: A good way to look at the problem of microforms is to look at it as an engineer would. In communication by radio, as you know, we have two kinds of wavelengths. We have the so-called audio, which is the long wavelength, with which we speak; the waves are many feet long. When we propagate these waves by means of transmission sets, the transmitter changes the frequency, reduces the wavelength, and sends it at a much shorter wavelength. This has the advantage that smaller equipment can be used and the wave can travel faster; in other words, it is a purely engineering device which changes the scale of the wave. In the same way in the visual field, we can take something which we can see with our eyes and we can reduce the scale to accomplish various purposes. Some of the advantages of reduction of scale Clapp has brought out, such as the fact that one could reproduce 10,000 pages for the cost of a single page when it is in microform; another advantage is that if one is going to move it in mechanical motion, an x, y, z, direction—such as in the AVCO system—the smaller the size of the record the faster the access time; in other words, the inertia of mechanical motion is lower. Thus you can get to any one of a million images, within a second or two, simply by reducing the inertia of the mechanical motion through the smaller size. There are other advantages not having anything to do with the user, such as the possible elimination of very expensive library storage as a trade-off against the very expensive equipment of bringing back these small scale images.

ROSE: Referring back to Vosper's comment about the statement in the report about resistance on the part of librarians; we meant nothing deprecating to librarians. We feel that the librarians have been resistant for some quite good reasons. On the point of esthetics, although we agree this is also a very important consideration, we considered it was beyond the scope of the paper.

We are dealing with something here in graphic storage that is in direct contrast to what we were discussing yesterday in terms of computers, in that we are dealing with information that is not quantified. Patrick's rules were very well stated; I think possibly the thing he neglected to mention is that there is nothing magic about computers and, by the same token, there is nothing magic about graphic storage systems. Computers will deal in a specific way, repetitively and complexly, with data that have been quantified, but there is nothing magic about the way the computer deals with it. This is the modern day answer to Babbage's calculating machine, so to speak. In the same way, in graphics we are not doing anything magic, we are only handling information that is not quantified. This is not to say it can't be quantified, but that it is not of its own nature quantified. And we are only handling this in a mechanical way. The *art* is in librarianship; it's only engineering techniques that we can offer to you.

We feel that the most important part of the problem is the systems problem, the intellectual problem; this is where the trade-offs come; this is where we have the competing alternatives. The things that the microform people and the machine people offer are in a very real way only pragmatic solutions. They are only tools. The real heart then of the paper with respect to the library is in the third section. The fourth section on system characteristics is really just an explanation of our definitions for the chart on page 115, the material in appendix B, and the application trends that you will find on pages 126 to 127.

MORIARITY: One of your needed research items should be to determine the genius of each of these graphic setups. How should one prepare copy for these? It's true, if you can copy the past for us, we'll be grateful; if you can copy it legibly, we'll be grateful to you! Most of the world's knowledge, however, is still to be developed, organized, and used—the past is over. What we want and what we should ask for is a way that the producers of information can prepare this material so that it will fit in adequately with this technology. I know the technology changes every day, and we must make our adjustments to it.

PATRICK: Several years ago I had a very small file to index and store. It consisted of styling drawings of automobiles, and we put them in aperture cards. The whole system broke down because we were unable to copy the aperture cards in case we wanted another copy or had other difficulties. A machine was announced only last month by IBM which will copy both the aperture—that is the film chips in the aperture—and add the coding in one magnificent operation.

WARHEIT: About 6 years ago the National Microfilm Association had a convention in Washington and just a few weeks ago it met in San Francisco. The contrast was one of the most surprising things to me; I have never seen such a change. What has happened, of course, is that for the first time microform is starting into the commercial areas. Because of the military aperture-card program there is now a large market and more people are designing machines, hopefully, production machines rather than expensive hand-built devices.

I once said if the typewriter were to be used only by librarians it never would have been invented, built, or designed. I think the same thing is happening here; the library market is small and the amount spent on research is small in comparison with that spent in the engineering-drawing market. But now it looks as if we will have a little copying machine and inexpensive readers, hopefully without hot spots. At least I am an optimist.

FUSSLER: I have been making bad and, hopefully, relatively good microfilm for roughly 33 years. It seems to me that the issues now and into the future can be stated relatively simply. First, there is the relative utility of the microform product to the consumer and to the institution that has to handle it and, secondly, there is the economics of the process. The economics has to include the cost to the readers as well as to the institution. It is idle to assume that we can force readers in scholarly institutions gladly to accept a product that has grossly inferior qualities at times in terms of its utility, or that libraries should accept a product and a process where the costs to the institution, in terms of production and storage and so forth, and the costs to the reader, which the institution may not be paying directly but will in the long run pay in one way or another, are higher.

MINDER: Perhaps if we want to become effective in getting some good copy and getting some improvements in the area of microreproduction, we should do it by way of standardization. Librarians should become important people in the standardization committees of the National Bureau of Standards and of the American Standards Association. Whoever is there to call the shots is going to set the standards. If we want high quality, we should be in with the standardization committees when they set the high quality. If we are not there, business takes over and makes it an economic matter, and we are going to continue to get a poor quality.

BECKER: I would like now to conclude this session. Thank you very much.

SECTION V

Output Printing

Output Printing for Library Mechanization

DAVID E. SPARKS, LAWRENCE H. BERUL, DAVID P. WAITE

Information Dynamics Corp.

Introduction

The library catalog in card form is a comparatively recent phenomenon and is the result of the growth of library work beyond the bounds of its technology. Sixty years ago the rate of growth of library collections had outpaced the ability of the book-form catalog to cope with the input; the solution to this problem was the card catalog. Technology, however, is catching up with the growth of library work, and through such devices as high-speed, graphic arts quality automatic typesetting, the library profession may one day be able to consider the reinstatement of the book-form catalog as a feasible working tool.

We should not, therefore, approach the mechanization of libraries with the idea that existing instrumentalities, such as the card catalog, are unalterable and basic. They are the current condition, but they, too, had their beginning and their historic reasons for being. Conditions which produced devices like the card catalog, however, are changing, and new methods are being developed to handle the data elements which form the library's basic data stores.

Among these methods automatic output printing has an important place. It is the interface between the internal activities of the library, necessitated by the automation goals, and the service goals established by the libraries themselves. That is, the fruits of library automation will be made available to the public through the technique of output printing. Automatic output printing is also, in the long run, the key to securing the broad base of user support which is necessary to make the cost of automation reasonable.

This paper attempts to review briefly the problems of output printing, especially automatic output printing, in the library. System approaches and equipment designs pertaining to automated printing subsystems of interest to libraries are undergoing an intense stage of competitive development at this time. The variety, scope, and complexity of output printing systems and equipments makes the task of presenting them most difficult. Nevertheless, a brief sketch can perhaps be given.

The first task is to state the products of the library output printing process. There follows a short review of various aspects of output printing of a technical or production nature interpreted in terms of the library's publication products. The next part of the paper deals with output printing equipment in some detail and is followed by a section devoted to matters of programming. A brief statement of the importance of systems engineering in the development of output printing systems is given in the section entitled "The Integrated Systems Approach." The last section presents some conclusions which, it is hoped, will help to put this part of the library mechanization problem in better perspective.

It is essential that the meaning of "automatic output printing" be clearly understood. Automatic output printing, or automatic typesetting, is not necessarily dependent upon the machine storage of the text to be printed. It is quite possible to drive an automatic typesetter from a computer, but the computer is not necessary to the typesetting operation. It is important to recognize, therefore, that automatic printing techniques are available for application to library publication problems, whether or not the data to be published are stored in particular library-based machinery. Thus the flexibility of automatic printing techniques allows for the existence of data stores that may or may not be machine-based. In this sense, automatic printing techniques are presently available to almost any library.

In a very real sense, a tape-operated typewriter is the simplest form of automatic output printing equipment. If we accept this broad definition it will be seen that even very modest libraries are capable of supporting some automatic output printing activities. Although the more complicated forms of output printing may be emphasized in this conference, in the broader view we should remember the popular forms of the art.

In this paper we have defined output printing in a mechanized library system as the full typographic composition of text with content, form, type style, leading, and white space as it will appear in copies produced for distribution. The concept of "the console" pertains to man-machine communications, for specially trained operators. Operator requirements at the console for printed text can be compromised to a large degree, according to engineering and machine design considerations. Publications for widespread use are quite a different matter, since consideration must be given to the basic reading habits, convenience, and comfort of the nonspecially trained users. Such a large audience is entirely outside of the library's training influence. The output printing products, therefore, will be a constant point of contact between the library and its patrons. Type composition used in these products must provide reading convenience as good as that provided by other printed material. Esthetic considerations are, of course, secondary to reading efficiency, but are not unimportant.

Library use of output printing techniques can range from the simple use of tape punching typewriters for simplification of internal routines to the development of completely automated typesetting processes. Between these two extremes are numerous possibilities for the use of special output printing machinery, shared data tapes, cooperatively operated service centers, and other arrangements.

Products for Library Output Printing

The published product of libraries can be considered from the viewpoint of the consumer for whom they are destined, of the use for which they are intended, and of the internal processes by which they are created. All of these points of view are valuable in a discussion of library publications.

Library Publications from the Point of View of the Consumer.—Printed output from the library is destined either for the patron (including the patron community as a whole) or for members of the library community. The primary bibliographical product of a library is produced therefore with the patron or the patron community in mind. This product can range from simple hand-typed memoranda delivered directly to the individual user to full-scale special bibliographies broadly disseminated to the patron community. The patron-directed output printing with which this paper is primarily concerned is that requiring preparation for and dissemination to a group of 25 or more. Such publications include, primarily, announcement lists and special bibliographies. Under conditions of automation considered in this conference, they could also include a record of results of mechanized searches accomplished by the library, both bibliographical listings and statistics of the searches performed.

In addition to this patron-directed output, the mechanized library would produce publications of interest to the library community, such as cumulated book catalogs, class schedules, and subject heading lists. Of interest to both patron and library communities would be the preparation of extensive periodical indexes. While beyond the scope of present techniques employed in most libraries, a product of this type could nevertheless be made available through the mechanized library as is amply demonstrated by the preparation of the *Index Medicus* by the National Library of Medicine.

Library Publications from the Point of View of Their Use.—Printed library publications have certain characteristics which are conditioned by their intended use. Publications of concern in the automated system are those intended for current awareness, retrospective search, or bibliographical control. Publications intended for the library community are mainly concerned with bibliographical control. Publications useful for retrospective search are of interest to both the library community and the patron.

Current awareness publications are used to make the patron rapidly aware of changes in the status of the library's holdings or capabilities. For this reason, these publications, e.g. announcement lists, news bulletins, etc., impose burdens such as publication deadlines on the library. In some instances these time factors may be quite important. On the other hand, current awareness publications are generally less demanding in terms of quality of typesetting, although this is not always the case.

These publications are usually widely distributed to the patron community, and, for this reason, the printing run or size of the edition may be quite large. It may, therefore, be of some importance to conserve paper (white space) in their production.

Library publications intended for retrospective search range from special bibliographies on specific subjects to complete catalogs of library holdings. The printing of this kind of material is less influenced by the necessity for rapid communication to the patron; nevertheless, they are not entirely free of the requirement for timely printing. On the other hand, publication of material useful for retrospective search usually involves considerable concern with the problems of updating and interfiling. In addition, these publications demand closer attention to the psychological factors involved in high-quality, graphic arts typesetting. There are two reasons for this situation: (1) the greater bulk of the publication requires more careful use of the printing space, (2) the material must be presented to the eye in such a way that the searcher does not experience fatigue. The size of the edition for publication of retrospective search material is certainly smaller than that required for current awareness material in terms of editorial units. In terms of total pages, it may be more. Nevertheless, it is difficult to predict what quantities of such materials could be distributed (i.e. what kind of a market could be found for them), if it were possible to produce them with mechanical means and in rapidly updated form.

Publications intended for bibliographical control are often not circulated outside of the library. There are, however, notable exceptions to this rule. This is especially so in the great national libraries whose published bibliographical tools rank among the most important of their products, e.g., the Library of Congress catalog cards, class schedules, and subject heading lists. Publications of the national libraries intended for bibliographical control have followed a slower production cadence than those publications intended for current awareness or retrospective search. That this has been so in the past does not necessarily mean that it is desirable. It indicates, rather, that in the allocation of their resources, the national libraries have often found it difficult to support the operations of editing and compilation necessary to produce these important publications for public consumption. The advent of automatic techniques in the internal operations of the national libraries, for which the encodement and manipulation of bibligraphical tools is essential, will certainly work profound changes in this situation. Cumulation and republication of subject heading lists and class schedules from the computer store will mean more frequent and larger printing runs of these publications.

The typographical quality of bibliographical tools is an important factor, since they are intended for continual visual lookup and as such are subject to the same rigorous demands for quality and legibility as are the publications intended for retrospective search. In addition, it has long been a practice to use typeface or type font changes as a means for conveying certain meanings within the data store. Typographical problems in printing these publications will therefore be greater than those encountered in printing other materials.

Library Publications from the Point of View of the Data Store.—In general, three types of data stores can be identified in libraries. These are: (1) the store of full text of the bibliographical items; (2) the store of bibliographical representations of the full-text items—the catalogs; (3) the store of auxiliary tools used for bibliographical control. Library output printing is concerned with these data stores, inasmuch as the text of the library's publications is composed from them. This is true whether the library is mechanized or not. For example, a special bibliography of materials in the library is usually composed from data represented in the library catalog. In the case of the mechanized library the connection between the machine-stored data stores and the output printing operation would be even more direct.

Library output from the full-text store is usually limited to photocopies or microfilm of the original text. It is evident that this will also be the case even in the automated library. The cost of storing the entire text of a large collection in computer memory would be prohibitive; however, smaller, special collections might be stored in some form of machine language. Storage of full text in other than the full-size printed form will probably be accomplished by some form of miniaturization. (See items 3, 17, and 38.)[27] The variety of normal and mechanized methods under development for creating, storing, and manipulating these microrecords is the subject of another paper at this conference.

Full-text items are represented in the library's files by bibliographical and lexical descriptions of which the catalog card is an example. Besides being the primary tools of bibliographical control, these files of bibliographical records are the major source of data contained in library publications. Announcement lists, special bibliographies, book-form catalogs, and special indexes all draw on these files for their data content. It is important to recognize the relationship between the output and the data store, since, from the systems point of view, desired characteristics of the published output must be represented in the mechanized data store. This relationship can be extended even further. Characteristics of the data desired in the final publication produced by the automated library should be represented in the input or original encodement phase. In this connection, recent advances in machine-interpretable input formats developed at the research library of the Air Force Cambridge Research Laboratories indicate that it is entirely possible to provide machine-interpretable input to large data stores with the simplest data processing equipment. (See item 33.) Decentralized libraries are thus able to provide a large centralized library with machine-interpretable[28] records of their accessions as a byproduct of their normal input processing techniques. In addition, these records can provide the decentralized library with the basic tool for partially mechanized publication or output printing if they hire a local service bureau for the data manipulation.

All libraries maintain auxiliary data stores as part of their apparatus for control of terminology, symbols, etc. Typical examples of such stores are classification schedules, subject heading lists, author authority files, etc. Complete or partial publication of these stores is not usually attempted except by the national libraries or other large libraries where publication is necessary for staff efficiency and by specialized libraries whose concentrated collections in specific subject areas require special expansions of class schedules or subject heading lists. Nevertheless, output printing from these auxiliary data stores may be an important part of the library's publication program. In the mechanized library, these stores will exist in machine-record form as an integral part of the data manipulation mechanism of the library.

Output printing of auxiliary data stores using high-speed typesetting techniques will be possible and, in the case of the national libraries, highly desirable. The attractive feature of this is, of course, the assurance of frequently updated editions of the class schedules and subject heading lists. Those who have experienced the difficulties of using class schedule supplements or who have had to invent subject headings because they could no longer tolerate delays can well appreciate the contribution which this kind of mechanized publication could make to the library profession.

Mechanization and automatic output printing of these bibliographical tools by the national libraries will result in additional benefits. Among these are the elimination of errors and redundancies in the subject heading lists and class schedules, and the development, by mechanical means, of a set of relationships between the subject heading lists and the class schedules. Further refinement of the language control tools in the direction of thesaurus building is also possible.

The development of a national standard for coding data in these auxiliary stores might encourage libraries that have developed specialized expansions of class schedules and subject heading lists to convert these to the standard encodement and thereby make available, through the national libraries, more powerful tools for bibliographical control. Other authority files now available at the national libraries (e.g. author authority files), but not offered to the public in printed form, could also be printed from a mechanized store.

In summary, output printing from the mechanized library will be directed to both patron and library communities. It will be produced for current awareness, retrospective search, and bibliographical control. It will be produced occasionally from the full-text store, but most frequently from

[27] This and similar references refer to items in the bibliography, p. 189.

[28] Machine interpretability implies a capability of mechanically identifying encoded data according to some instruction, in addition to readability.

the store which we, today, commonly call the main catalog, and, in the case of the national libraries, from the auxiliary store the development of which is so important to libraries throughout the land.

Aspects of Library Output Printing

To identify necessary, desirable, or potential output printing tasks of the library is only the first step. These tasks must each be specified both in terms of publication objectives and technical characteristics. It is especially important to accommodate the publication objectives to the constraints of the mechanized process.

Those who design output printing systems for libraries must have detailed statements of output printing requirements on the basis of which decisions can be made concerning processes and machinery to be used. It is not sufficient to state these requirements in general terms; they must be stated in terms of specifications meaningful for system planning. These specifications might include (1) the size of the text being printed stated in terms of the total number of characters in the text; (2) the number of copies to be printed in the printing run; (3) the frequency with which the item is published or republished; (4) the number of character sets or fonts required to set type for the publication; (5) the psychological or semantic considerations involved in arranging the type on the page; (6) the manipulations required to arrange the pieces of the text for publication. These specifications must be considered for each of the identified library publications.

Characteristics of Library Publications.— *Text Packages.* Library publications have some characteristics which are unique. The most outstanding characteristic is that most of the textual content of library publications is piecemeal. That is to say, it is rare to find library publications consisting of continuous discourse; in general they are assembled from small packages of text. In addition, a small package of text may be repeatedly printed in different contexts. Preparation of a publication will very often require manipulation and assemblage of these basic data packages. This is of interest because some of the mechanized techniques for output printing include, or are closely connected with, routines for file manipulation. With respect to file manipulation, library output printing can be divided roughly into two categories: publications whose text is assembled from data packages, and publications whose text is composed of discrete words ("word" is here used in the sense of any typographic combination).

Publications produced by assembling data packages are announcement lists, special bibliographies, book catalogs, and lists of serial holdings. Publications consisting of text composed from discrete words include library catalog cards, subject heading lists, classification schedules, statistical reports, and authority files.

Prepublication manipulation of data, or data packages, is a most important aspect of library output printing. For this reason, the distinction between publications assembled from standard data packages and those composed of individual "words," or those assembled from data packages with inserted "words," may be important in designing mechanized printing operations for the library.

Size of Text. Another characteristic of library publications is the extreme range of size—from a single hard copy of a bibliographical reference to a large edition printing of a multivolume catalog. In general, publications of the shortest length are those produced as a result of direct patron requests. Such publications are the result of individual search and retrieval operations and consist of a few references, a limited special bibliography, or they satisfy requests for copies of individual catalog cards. Other limited publications will result when announcement lists, news bulletins, and library cards are produced at high frequency. Publications having fewer than 100,000 characters would be normal for the major portion of the output printing of the library, e.g. announcement lists, special bibliographies of more than trivial length, statistical reports, and supplements updating larger publications. A printing load of up to a million characters can be expected for special bibliographies, book catalogs, card catalogs, subject heading lists, class schedules, authority files, and serial holdings lists; indeed some of these publications may require many millions of characters.

Size of Printing Run. Output from the library may be a single copy distributed to a single patron, or it may be disseminated broadcast to a very large patron population. In considering the size of the printing run two general levels of activity can be distinguished. Output is often required in trival amounts of from one to five copies, and this kind of publication will undoubtedly be produced by other than formal printing techniques. Representative publications of this kind include individual search results consisting of a few references, small special bibliographies, and the printout obtainable from the monitor console or high-speed mechanical printer. In contrast, library publications can have an edition size of many thousands. Examples of such publications are announcement lists, special bibliographies, statistical reports, authority files, serial holdings lists, class schedules, and subject heading lists. Production of catalog cards by the national libraries also represents a substantial publication effort. The more frequent and timely publications will usually be distributed to a wide audience of patrons and other libraries and will require larger editions.

Frequency of Publication. Output printing from the library can be classed into three groups according to the frequency of publication. There are publications which are produced periodically, occasionally, and upon demand. Periodic publications range in frequency from daily publications to those that are produced every 5 to 10 years. Most usual are the monthly publications such as announcement lists, newsletters, and supplements to larger publications. Quarterly publications include, or could include, special bibliographies, cumulations of book catalogs, cumulations of subject heading list additions and changes, cumulations of classification schedule additions and changes, and cumulations of serial holdings lists. Cumulation of many library publications would be on an annual basis. Occasional publications are those which require long planning and editorial work, such as cumulated book-form editions of the library's catalogs, classification schedules, subject heading lists, and authority files. Demand publications are, by definition, those which result from patron requests. As such, they are somewhat beyond the scope of the general printing activity of the library, except that printing activity associated with a monitor console or a high-speed mechanical printer. Certain frequently requested items in the library, however, might be preprinted in advance of patron requests if demand for such items could be predicted from computer analysis of activity statistics.

Response Time. The allotted publication time is of importance in planning output printing operations. Response time is defined as the time between the editorial "go" and the time the publication must be distributed to the consumer. Response time varies from the immediate (face-to-face patron requests) to moderately long range (quinquennial cumulations of book catalogs). The most important requirement for speed comes under the pressure of a specific patron request for search and retrieval. Demand for rapid response in this situation will influence the choice of output printing techniques. Immediate response printout is a function of the monitor console; slightly longer response times for patron requests can satisfactorily be handled through the use of standard typewriting equipment or mechanical output printers. The console display in a fully mechanized, man-machine system will play an important role in the on-line fulfillment of patron demand. However, as will be mentioned later, the requirements of the console display are at some variance with the demands for library output printing.

Time demands for the majority of library publications are perhaps less exacting. Current awareness publications must, of course, be produced under tight publication deadlines. There is also the reasonable demand for frequent and regular updating of bibliographical tools, although these are influenced more by the demand of interfiling and textual accuracy than they are by the time factor. The bulk of library output printing generally requires a response time ranging from 1 to 2 weeks for monthly publications to 1 to 2 months for annual cumulations. Response times for large cumulations will be significantly shortened through the mechanization of the prepublication manipulation activities necessary for the preparation of the text. Response time is not independent of the bulk of the text to be published. As the bulk of the text increases, it becomes increasingly necessary to seek high-speed composition and printing techniques in order to reduce the time between the editorial closing date and the actual press work. Thus, even the large, long-range publications are subject to the need for reduction of the response time.

Quality of Printing. Quality of library output printing is again a factor subject to a range of requirements. In the case of the rapid response to a patron request, the quality of the output printing may be sacrificed for speed. Publications less affected by the demand for speed can be produced with more attention to typographic quality.

Typographic quality is defined by several considerations which are a combination of technical and psychological factors. Most important of these are the choice of typeface, the point size, the use of upper and lowercase, the alternation of typefaces, and the skillful use of the ratio of text to white space surrounding it. Studies concerning these factors and many others are summarized in in a recent report to the U.S. Navy Bureau of Ships (see item 39).

A general characteristic of library publications which affects the demand for typographic quality is the piecemeal nature of the text. Perusal of bibliographical publications consists primarily in a scanning or searching activity, rather than a reading activity. For this reason, careful attention must be paid to providing those visual cues which can assist this searching activity. Appropriate use of type size, position, boldface, italics, underscoring, and white space can achieve this result.

These psychological considerations apply to all the output printing products which we have identified. They are especially important in those publications which mass many different information elements on a typographical field or page. (See fig. 25.)

The quality of output printing is also related to other factors. The number of copies in a printing run often influences choice of type quality in the publication. Short-run publications (few copies) will not be composed with the care given to publications intended for wider distribution. The range of edition size mentioned above indicates that the ability to make a choice of type quality must be built into the library's output printing system.

Whether a library publication is ephemeral or is intended to have some degree of permanence will also influence the choice of type quality. A publication intended for long use will require care in typesetting just as it requires care in editorial preparation.

Finally, library publications, because they are composed largely of bibliographical citations and only infrequently of running text, must make skillful use of typesetting as an aid both to their composition and to their use.

Character Sets. Equipment to fill these needs for quality output ranges from single-spacing, single-case machines to the most complicated typesetting devices. The character sets required range from single fonts of single case to large multiple-font sets.

The on-line patron request can probably be met by single font, high-speed machines, although the cumulated results of search and retrieval activities may be produced using a higher quality typographic output.

Outside the patron-response situation, however, library output printing, it would seem, requires more sophistication in its use and display of typographic characters than is available in single font equipments.

Special attention must be given to the library practice of using changes in typefaces or of type fonts to convey certain meanings. This practice is followed in the Library of Congress class schedules and, to some extent, on the Library of Congress catalog card. Mechanization of typesetting for such publications must take these semantic uses of typography into consideration.

Generally speaking, the more timely the output printing product of the library, the less demanding the requirements for a variety of character sets or fonts. Widely disseminated monthly announcement lists may be produced with a simple monospacing, two-case, tape-driven typewriter. On the other hand, a carefully edited special bibliography may require several type fonts, including even non-Roman alphabets. The larger and more permanent publications require more careful selection of character sets or fonts.

It is extremely difficult to establish a range of character sets for library output printing. Nevertheless, the following observations can be made: (1) output printing for immediate response to patron requests can probably be handled with the characters on presently available computer output equipments; (2) publication of the generally disseminated library products and of the frequently cumulated publications can probably be accomplished with a character set of 300 elements; (3) larger, more comprehensive, and more carefully edited publications whose content requires representation of a wider range of language

FIGURE 25. *Typographic quality of various output devices.*

and composition elements will probably require character sets of 500 characters or more.

Production Aspects of Output Printing.—In applying methods of type composition to specific printing problems, several general factors must be considered. These relate to the volume of composition, the number of type fonts and faces, the size of the printing run, the effect of type density or type productivity, and problems of production such as work scheduling and load peaking.

Quantity of Text. Type composing systems must take into consideration the total amount of text to be set in any given production run, since this is a governing factor in the specification of the size, capacity, and speed of the automatic type composing equipment to be employed. In this connection, one must be aware of the maximum amount of text that will be encountered in a production run and also of the range of the text sizes and the average composition task.

Number of Fonts. The volume of composition and the quantity of type required for the particular application of typesetting determines the number of type fonts or typefaces required. The simplest requirement for typeface is continuous English text. If the text is to be set in only one face, a single alphabet is sufficient. Provision of special fonts, however, such as italics, boldface, small capitals, etc., demand additional capabilities in the typesetting system. Typeface changes normally used in library output printing of, say, the Library of Congress, would require the extensive multiplication of character sets. Changes of type font are involved when changes of point size or typeface are indicated and when text is set in non-Roman alphabets. These changes put added burdens on the system and make it more difficult to establish the basic character set.

Type Density or Type Productivity. The matter of type density or type productivity is of interest to the designer of typographic composition systems. Type density is measured in the number of characters per square inch of the printed page and is a function of the point size of the type, the characteristics of the font, and the boldness of the typeface. Other factors affecting type density are the intercharacter spaces required for right-hand margin justification and interlinear spacing or leading. Type density is a measure of the efficiency of typesetting in making use of printing space.

Type density raises psychological considerations. A text which is printed in densely packed type may be difficult to read; on the other hand, text which is dispersed by too much white space between characters and by too much leading between lines can also be difficult to read. Optimum type density is a matter of balancing the demands of the text to be set with measures of readability. An example of the effect of type density is shown in figure 26 which shows the same text as composed by a computer output printer and by a high-quality type composing machine. The example shown was taken from the Defense Industrial Supply Center (DISC) *Catalog*. The upper portion was produced from a computer printout which has been reduced photographically to achieve a type density of 17 characters per inch. The lower portion shows the same text in News Gothic typeface reduced to 7.5 points with an average type density of 26 characters per inch. A sample of 7½ pages of this material composed in the News Gothic face was found to contain the same text which, in the DISC *Catalog* computer printout and format, required 20½ pages.[29]

Operational Problems. Production problems in the typesetting operation include estimating total quantity of text to be set in a given period and the periodical accumulation of workload. Obviously, the productivity or density of the typesetting will influence the total quantity of the text, considered in units of output pages, and thus the cost of the printing run. Where large quantities of text are involved, close attention to type density may result in savings of many thousands of dollars. Careful choice of typeface and judicious use of typesetting skills must be considered integrally with the other factors in the printing process. The possible accumulation of the workload at periodic intervals is of importance in estimating the cost of an output printing operation. Peak workloads may demand high capacity and expensive typesetting systems that cannot be justified during slack periods. When this is the case the solution may lie in the sharing of equipments by several libraries or the

[29] From this study carried out by the U.S. Navy Publications and Printing Service, it was estimated that savings in printing costs alone for each republication of the DISC *Catalog* would amount to $750,000 if graphic arts quality typography were employed.

UNIT OF ISSUE: EACH (EA), UNLESS OTHERWISE INDICATED

INDEX NO.	FEDERAL STOCK NO.	SIZE, IN.	MESH	WIRE DIA IN.	MFR CODE NO.	MFR PART NO.
	RECTANGULAR					
• 5	5335-550-4163	1.84 LG X 0.27 W	60 X 60	.0075	76050	MC2815
10	5335-685-6749	1.32 LG X 0.60 W	20 X 20	.016	02734	8412881-2
15	5335-685-6748	1.40 LG X 1 W	20 X 20	.016	02734	8412881-1
20	5335-392-8598	3-15/32 LG X 2-13/32 W	18 X 18	.011	--	--
• 25	5335-524-8704	6.406 LG X 2.656 W	60 X 60	.007	77200	02-570
• 27	5335-833-5453	6-7/8 LG X 2-3/4 W	20 X 20	.015	09975	DRA45956-1
• 29	5335-833-5452	14-3/8 LG X 5 W	20 X 20	.015	09975	DRA45956-2
• 30	5335-623-5053	3-3/8 LG X 3-5/16 W	8 X 8	.025	--	--
• 35	5335-425-1367	60 LG X 48 W	--	.192	--	--
	ROUND					
40	5335-531-0059	.36 DIA	--	.0135	--	--
45	5335-201-1818	.500 DIA	60 X 60	.010	82267	3-151-36
50	5335-584-2145	.562 DIA	100 X 90	.0045	45413	100143
55	5335-664-3780	.749 DIA	250 X 250	.0016	98939	A21129
60	5335-721-7294	.760 DIA	100 X 100	.003	--	--
65.	5335-637-2877	.985--1.000 DIA	30 X 30	.013	59875	TV95131
• 70	5335-264-4142	.980--1.010 DIA	28 X 28	.010	--	--
• 72	5335-827-1243	1.226 DIA	8 X 14	--	00000	9147892-1
• 75	5335-579-9908	1.875 DIA	80 X 80	.007	87991	132118-00
80	5335-575-4311	2-1/2 DIA	84 X 84	--	00975	A-157
• 82	5335-819-7288	7-1/2 DIA	2 X 2	.063	01066	7212881P2
85	5335-806-7747	15.44 DIA	60 X 60	.0045	08203	59-205-1216

(a)

Unit of Issue: Each (EA), unless otherwise indicated

Index No.	Federal Stock No.	Size (In.)	Mesh	Wire Dia (In.)	Mfr Code No.	Mfr Part No.
RECTANGULAR						
•5	5335-550-4163	1.84 Lg X 0.27 W	60X60	.0075	76050	MC2815
10	5335-685-6749	1.32 Lg X 0.60 W	20X20	.016	02734	8412881-2
15	5335-685-6748	1.40 Lg X 1 W	20X20	.016	02734	8412881-1
20	5335-392-8598	3-15/32 Lg X 2-13/32 W	18X18	.011		
•25	5335-524-8704	6.406 Lg X 2.656 W	60X60	.007	77200	02-570
•27	5335-833-5453	6-7/8 Lg X 2-3/4 W	20X20	.015	09975	DRA45956-1
•29	5335-833-5452	14-3/8 Lg X 5 W	20X20	.015	09975	DRA45956-2
•30	5335-623-5053	3-3/8 Lg X 3-5/16 W	8X8	.025		
•35	5335-425-1367	60 Lg X 46 W		.192		
ROUND						
40	5335-531-0059	.36 Dia		.0135		
45	5335-201-1818	.500 Dia	60X60	.010	82267	3-151-36
50	5335-584-2145	.562 Dia	100X90	.0045	45413	100143
55	5335-664-3780	.749 Dia	250X250	.0016	98939	A21129
60	5335-721-7294	.760 Dia	100X100	.003		
65	5335-637-2877	.985--1.000 Dia	30X30	.013	59875	TV95131
•70	5335-264-4142	.980--1.010 Dia	28X28	.010		
•72	5335-827-1243	1.226 Dia	8X14		00000	9147892-1
•75	5335-579-9908	1.875 Dia	80X80	.007	87991	132118-00
80	5335-575-4311	2-1/2 Dia	84X84		00975	A-157
•82	5335-819-7288	7-1/2 Dia	2X2	.063	01066	7212881P2
85	5335-806-7747	15.44 Dia	60X60	.0045	08203	59-205-1216

(b)

FIGURE 26. *Relative typographic efficiency (a) computer output printer (b) composing machine.*

use of several types of equipment for several levels of operation, or both.

From the point of view of production, the cost of operating a typesetting system is a function of the production load. Figure 27 illustrates the effect of the production load on the unit cost of operating typographic composing machines. The particular machines selected for the comparison were chosen to cover the range from a relatively slow device to a highly productive machine. The unit costs represented were computed from manufacturers' ratings or from engineering estimates as applicable; they are rough approximations, not intended as an evaluation of the equipments themselves. Nevertheless, the effect of load on composition costs in the utilization of a highly productive output printer is clearly evidenced.

When type productivity is taken into consideration as an element of production cost the question of the efficiency of the mechanical computer output printer is raised. Figure 28 shows how the total annual cost for publishing a hypothetical announcement journal (including a number of indexes) is related to the number of copies. Again, rough approximations or estimated figures are used. It can be concluded from this illustration that the efficient use of type in the composition process is a major element in all printing cost considerations and becomes more important with the size of the production run.

Technical Aspects of Output Printing.—In studying the problem of output printing for the mechanized library, library administrators should be aware of certain details of type composition, beyond production problems, which have a direct bearing on cost considerations. It is not possible to enter here into a detailed description of these factors which are the basic subject matter of a trade with a long history and ample documentation.

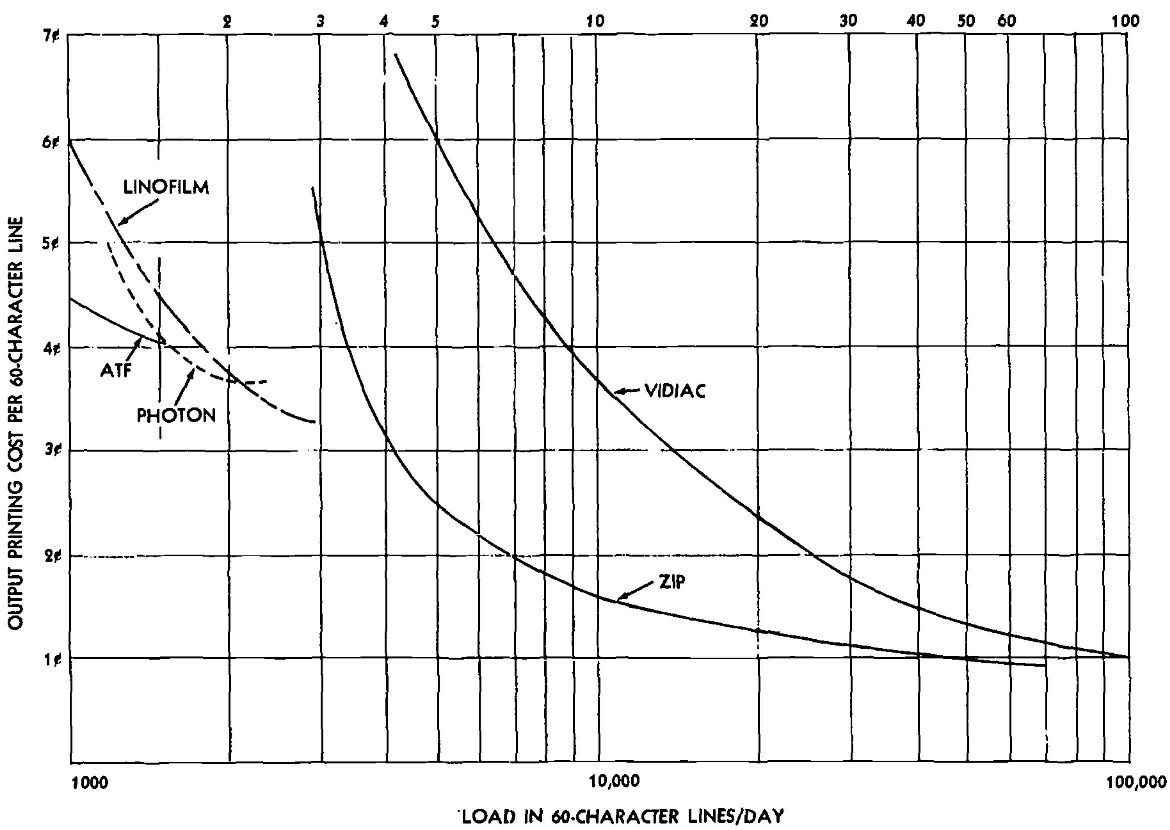


FIGURE 27. *Total unit cost vs. load—tape operated photocomposing machines.*

735-898 O—64——12

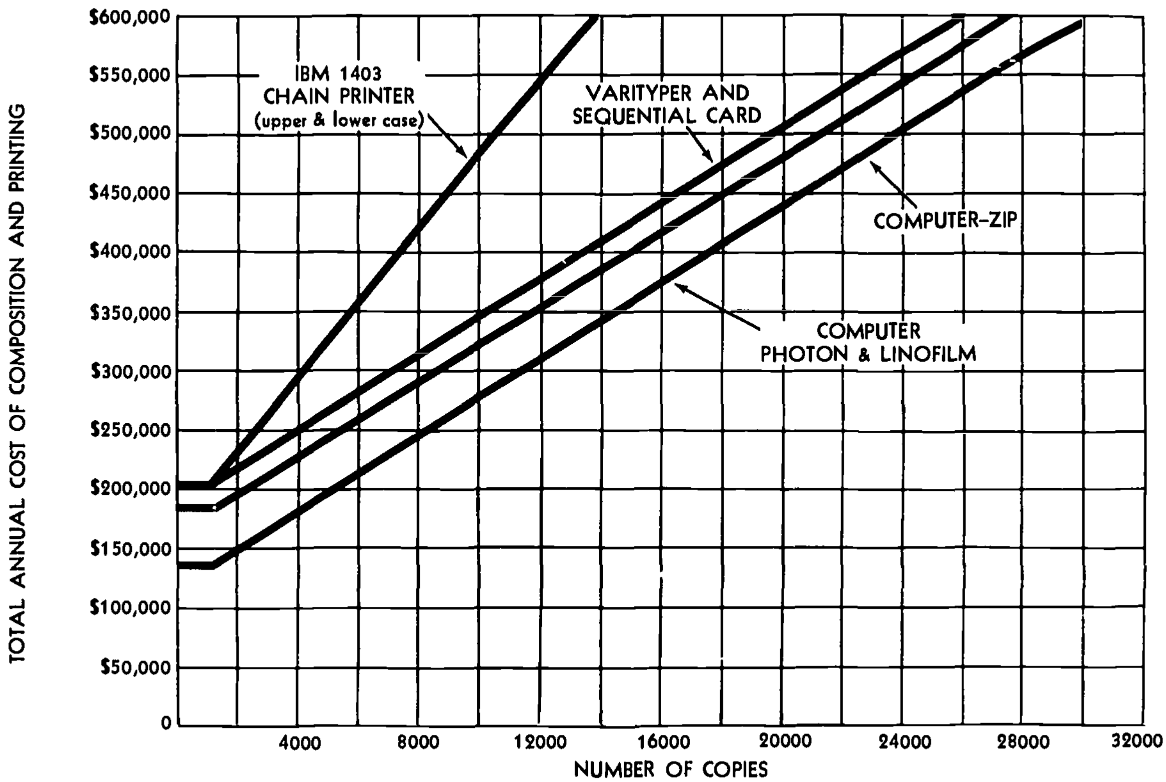


FIGURE 28. *Total annual cost vs. number of copies.*

They are mentioned here briefly to make the reader aware that they play a part in the output printing problem.

Horizontal Justification. High-quality printing, such as Linotype, Monotype, Linofilm, etc., produces copy with even right-hand margins. The process of producing the even right-hand margin is known as horizontal justification. It is usually accomplished by increasing the spaces between words and occasionally between appropriate pairs of adjacent characters.

Special keyboards are employed to cope with the problem of horizontal justification. These keyboards (illustrated in fig. 29) keep track of the remaining space left in the line. Since different typefaces all have different set, or width values, it is necessary to look up the width value of each character in the selected font and subtract that width value from the total remaining space in the line being set. Individual width cards or magazines which are inserted in the keyboard device are supplied for each type font. The keyboard device also keeps track of the number of interword spaces which have been inserted in the line. These word spaces may be expanded from a minimum to a maximum value thereby providing a justification range of some latitude for the keyboard operator.

The process of width-card use, line length, and justification can be handled internally in a computer. The complexity and cost of initial keyboarding is thereby greatly reduced. The programming problems to accomplish these tasks in the computer are a challenge, however, which have been attacked and solved to some extent already by a few workers in the field. Some doubt has been cast upon the value of horizontal justification by psychological studies designed to test its effect on reading efficiency. Nevertheless, it has a psychological value rooted in long tradition.

Currently, horizontal justification is still used in almost all major printing activities. The increase in the use of cold type composition methods, however, has brought about a tendency to disregard horizontal justification and substantial publications may be found today that disregard it.

Horizontal justification is concerned with the width of the line of type. Line width is of particular importance in library printing since so many of the publications involved are composed of textual packages of specific character lengths. These packages of text are usually cumulated and assembled in columns which are substantially smaller than the full line width of the page. A feature of horizontal justification is the fact that the difficulties of justification increase as the line width decreases. This is so, because a shorter line has fewer letters, fewer words, and fewer spaces between words which can be used to expand or contract the line for the justification process.

Hyphenation and Reformatting. The problems of hyphenation and reformatting are closely related to the justification and line-width problems. Like justification, hyphenation is more difficult in text set in short lines. It has been shown statistically that text in short lines will have a higher incidence of hyphenation problems than text set in the normal page width.

Reformatting is related to hyphenation since it involves the repositioning of characters and words in lines different from the original composition. Reformatting may be a serious problem in type composition for library output printing, because much of the text of library output printing is assembled from manipulated data packages. During the course of such manipulation, it may be necessary to reformat whole lines of type. An example

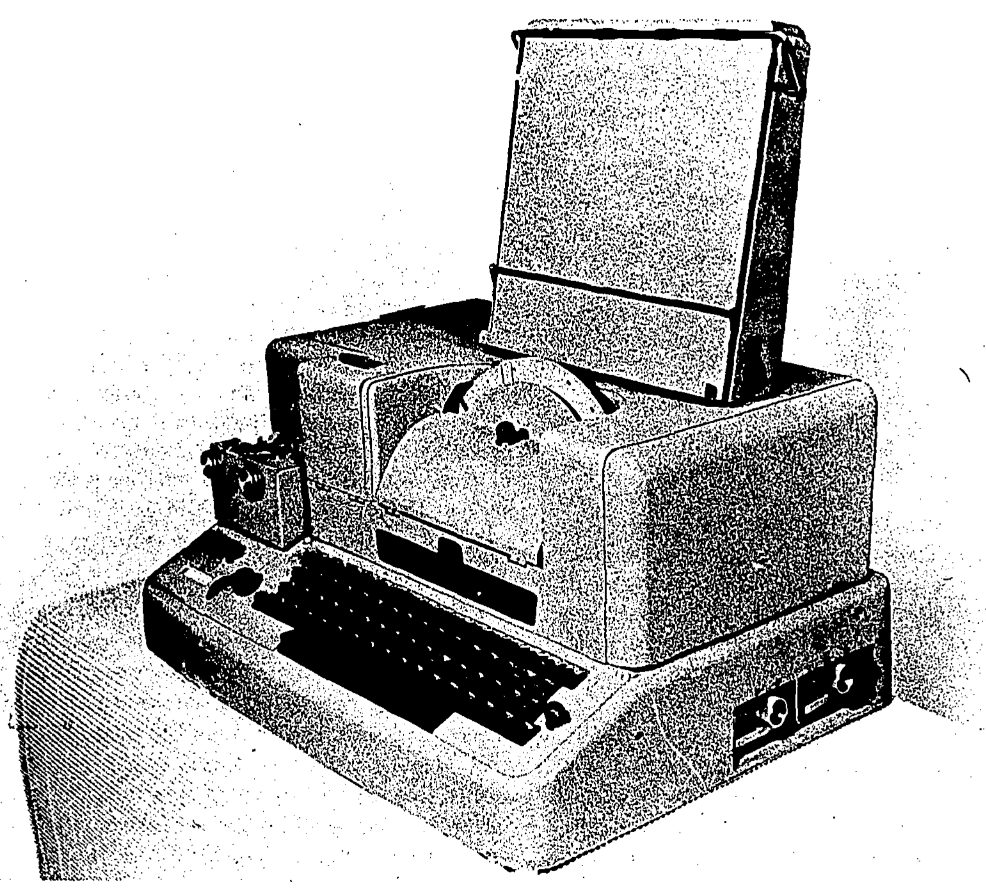


FIGURE 29. *Teletypesetter Universal keyboard with horizontal justification capability.*

of such reformatting would be the typing of parts of titles as added entries where the words in the body of the library card are spread over two lines of type but are to be laid out in one line of type in the added entry heading.

Leading and Vertical Justification. The space between lines of type can be varied to increase legibility or to conserve page space. This variation of interlinear space in type composition is called "leading," from the insertion of thin strips of lead between lines of type in the manual typesetting operation. Leading of type is an important factor in the efficiency with which the typesetting operation is performed. Leading is also used in the composition of type in columnar form when it is desired to make two columns even at the top and bottom. This is known as vertical justification or column balancing. Since much of library output printing will involve composition of material in column widths, the problem of vertical justification must be taken into consideration.

Other Technical Problems. Minor type composition techniques include the use of margins and gutters in page composition, the control of "rivers" in text, and the patterning of typeface. The latter is a technique of emphasizing words by the use of bold or italic faces so that they are easily visible in text. Patterning could well be used in computer composed text as a means for emphasizing search points without resorting to extensive reformatting.

Automation Aspects of Output Printing.— Certain aspects of automatic type composition which have a special bearing on library printing problems should be mentioned. These concern the relationship of the input data to the composition process. Input data for automatic type composition processes usually consist of some form of punched paper encodement. The generation, configuration, and manipulation of these encodements can have some importance to the mechanization of libraries.

Automatic Composition Without Computers. Libraries may have output printing needs, even though they do not have access to, or cannot afford to utilize a computer. These libraries may prepare machine-interpretable records of their cataloging and indexing operations for possible future use, or for input to some other mechanized library or library processing center. Once this cataloging information has been captured in machine-interpretable form, it can be utilized by less sophisticated but nevertheless potent devices which can automatically or semiautomatically assist in the output printing function. A discussion of these noncomputerized uses of automatic type composition will be given in the following section entitled "Output Printing Equipment." Thus, mechanized output printing is not limited to libraries that can afford investment in computing machinery. On the contrary, many systems of automatic typesetting or automatic text composition are available at modest cost. In addition, the development of computer service centers throughout the country is creating opportunities for libraries to make use of off-line computer processing techniques.

Code Conversion. In order to operate a composing machine or output printer from paper or magnetic tape, it is necessary that the tape be in the code and format required by the particular composing machine. If the paper tape generated at the input keyboard is not in the proper code and format for the output printer, there are two possibilities. Where the data are stored in a computer, the computer can convert its internal code into the proper code and format required by the output printer. If the output printer is operated from paper tape, the tape can be produced either on-line with the computer or off-line as a separate operation. If a large scale computer, such as an IBM 7090, is used as the central processor it will probably be inefficient to convert to paper tape on-line. On the other hand, smaller scale systems such as the RCA 301 and IBM 1401 can economically be used on-line to format paper tape for the output printer.

Typographic Control Functions. A number of controls must be encoded into the tape that will operate the output printing device. These codes will instruct the printing device to select characters, shift, change fonts, and change lens. Codes will also be used for formatting by indicating end of line, relative spacing, quadding, number of interword spaces, line deficit, leading, and character-width values. These control functions can be inserted during the initial keyboarding or as part of the output edit programming subroutine.

Obviously, if the output printing is to be produced from information contained in the computer store, most of these control function codes can be generated by the computer. On the other hand, if a library wishes to utilize high-quality output printing devices, such as Photon or Linofilm, without a computer, its input keyboarding functions will be considerably more complicated and must take these control functions into consideration. It is possible, however, to serve both functions, i.e., providing input to a computer for information retrieval and also providing a tape for the operation of a tape-operated composing machine.[30]

Output Printing Equipment

Output printing devices may be divided into three categories: (1) those operated by relatively low-speed perforated storage media such as paper tape and punched cards; (2) graphic storage media exemplified by the sequential card camera; and (3) high-speed magnetic-tape-operated devices including mechanical and nonmechanical computer printers and graphic arts quality composing machines.

Paper-Tape-Operated Composing Machines.—*General.* Paper-tape-operated composing machines operate at a speed of 5 to 20 characters per second (cps). The typographic quality of these machines ranges from that of the standard electric typewriter to that of the high-quality composing machine such as Linotype or Photon. A single tape-operated composing machine may range in cost from as little as $2,000 to as much as $50,000, the difference in cost being essentially a matter of typographic quality and flexibility.

Three types of paper-tape-operated composing machines will be discussed: ribbon impression machines, hot-lead typesetting machines, and photocomposing devices. They will be presented in the order of increasing output speed and mechanical complexity.

As was mentioned above, devices associated with these machines produce punched paper tape encoding not only the characters of the text but also instructions to the machine which control justification, typeface, etc. This perforated paper tape may be generated directly by an input keyboard, such as a Flexowriter, or as an output product of a computer. The necessary control codes for type style changes and format can be inserted in the tape, either at the manual keyboard stage, or automatically by the computer. We should therefore think of these paper-tape-operated devices, about to be described, both as a type of computer output printer and also as an automatic printing device for the library which does not require a computer.

Ribbon Impression Composing Machines. This type of device consists of a standard typewriter keyboard with a paper-tape reader and a paper-tape punch. (Fig. 30.) The paper-tape punch produces machine-readable tape while at the same time the type basket produces standard typewritten copy. The paper tape may also be read back into the paper-tape reader to automatically produce typewritten copy. There are usually 44 character keys making 88 characters available by means of upper and lowercase shift. Such devices may have utility in libraries that wish to record their cataloging information in machine-readable form. The paper-tape byproduct of the initial typing may be reused to produce as many copies as desired by inserting it in the paper-tape reader which automatically operates the typewriter. The typing speed, when tape operated, ranges from 10 to 15 characters per second. Some of these devices have what are referred to as "programmatic" features so that the machine can be programmed to automatically insert codes or standard portions of text, tabulate, return carriage, skip lines, punch or not punch a secondary tape as desired, or not to print information which is to be punched but not typed.

A feature which is available on only a few machines of this type is proportional spacing. Proportional spacing produces a higher degree of typographic quality because the relative width of each character is taken into account in the design of the typeface and of the carriage escapement. It consequently improves the appearance of the page as well as reduces the total number of pages by a factor of 10 to 15 percent.

[30] As an example of composing on the input side of a computer, Information Dynamics Corp., of Wakefield, Mass., developed a tape-operated system for photocomposing the NASA journal, *Scientific and Technical Aerospace Reports* (STAR). The system utilizes a modified LCC–S keyboard to produce a paper tape which drives a Photon photocomposing machine. The tape is used also as input to an IBM 1401 computer for information storage and retrieval.

The smaller libraries which plan to encode their cataloging information might utilize one of the ribbon-impression, tape-operated composing machines for producing sets of catalog cards, accession lists, or other products previously described.

A special purpose device has been developed to manipulate a machine-interpretable paper-tape record of a Library of Congress catalog card and to produce an expanded tape which can be used to produce sets of diversely headed catalog cards automatically. (See item 33.) This expanded output tape is used to operate one or more tape typewriters. The name of the device is the Itek Crossfiler, developed by Itek Corp., of Lexington, Mass., for the Air Force Cambridge Research Laboratories.

Examples of ribbon impression composing machines which can be used as output printers as well as input keyboard devices (assuming that code compatibility and format problems can adequately be solved) are:

1. Friden Flexowriter
2. Friden LCC-S Justowriter
3. Friden Justowriter
4. Remington Rand Synchrotape
5. Smith Corona Typetronic
6. IBM 870 Document Writer
7. Dura Mach-10
8. Invac TTR-200
9. Invac TTR-100
10. Invac PK-144

These devices range in cost from approximately $2,000 to $5,000.

Hot-Lead Typesetting Machines. The hot-lead composing machines developed by Mergenthaler and Lanston in the latter part of the 19th century were the first practical mechanized composing systems. They produce copy of graphic arts qual-

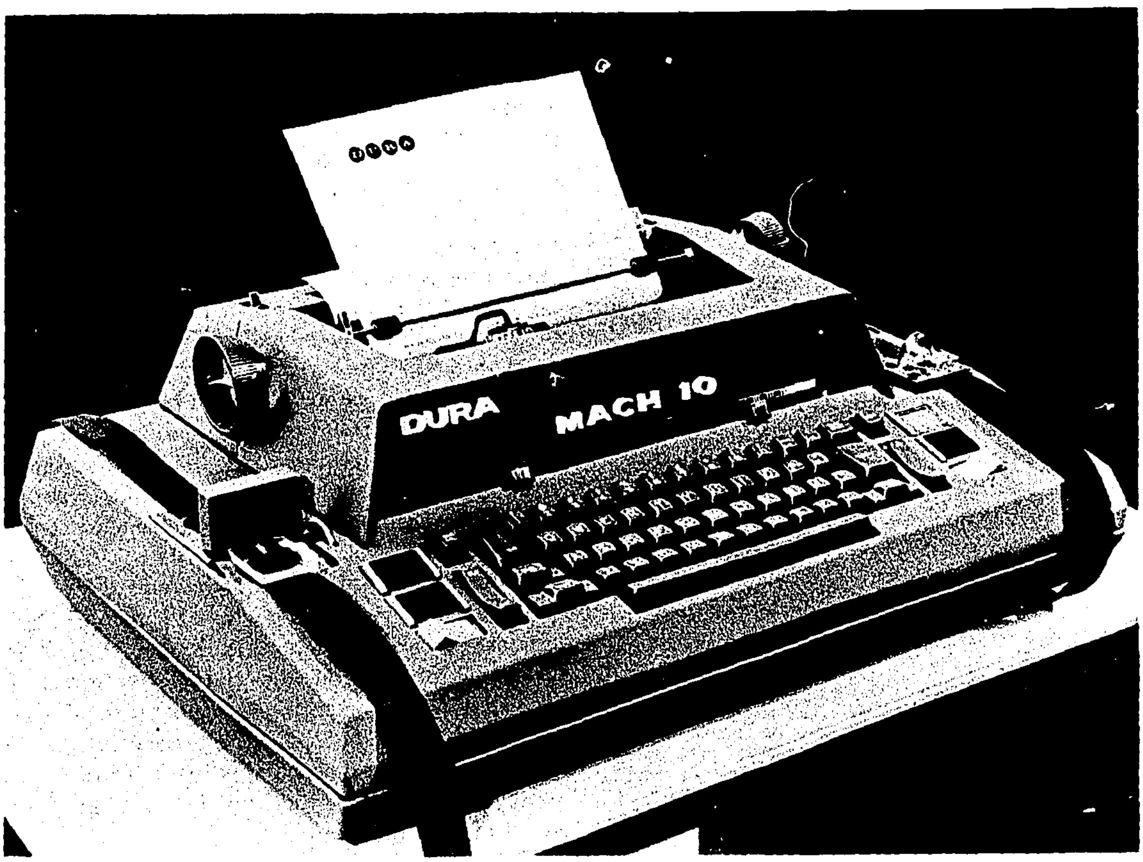


FIGURE 30. *The Dura Mach 10—a ribbon impression composing machine with paper-tape punch and reader.*

ity and since their development have been the standard of the printing industry. These lead-casting machines are equipped with a supply of brass matrices for each individual character in the font. When a key is struck on the keyboard or actuated by paper tape, the appropriate matrix is released from its container (magazine) into an assembler until a line is completed. At that time molten metal pours onto the assembled line of brass matrices and in this manner a line of type is cast and then ejected from the mold. The individual brass matrices are then returned by a distribution system to their magazines so they may be used again. High-speed slugcasting machines are capable of producing 375 to 900 newspaper-column-width lines per hour. The new Linotype Elektron (shown in fig. 31) is the fastest slugcasting machine made; it operates at the maximum speed of 900 newspaper-column-width lines per hour. The Intertype Monarch can deliver up to 840 such lines per hour. Converted into characters per second, the maximum effective tape operating speeds of even the fastest hot-lead slugcasting machines are still a very modest 10 to 13 characters per second. The usual code format for tape operation of a slugcasting machine such as Linotype is the 6-bit Teletypesetter (TTS) code structure. Slugcasting has been with us for over 70 years; the tape operation of such machines since 1927. As a result, a wealth of technical and descriptive material is available that explains this technique in detail. For example, see item 1.

The Monotype developed by Lanston comprises separate keyboard and casting units. The keyboard resembles a giant typewriter and carries a vertical graduated drum and a horizontal scale indicating the remaining space left in the line being set. The principal difference between the operation of a Monotype and a slugcasting machine such as Linotype and Intertype is that the Monotype casts single pieces of type rather than entire lines. This makes it easier to perform simple corrections and it is therefore frequently utilized in difficult composition work such as scientific materials involving mathematical equations. Monotype composition is generally more expensive than Linotype and is slower. A second difficulty is that the punched paper tape which operates the Monotype caster is incompatible with most computing equipment since it uses a 31-bit code and a tape width of 3¾ inches, compared to the more widely used ⅞ or 1 inch paper tapes. The Monotype matrix can hold up to 2[illegible] different characters.

It is important to recognize that the paper tapes used to operate hot-lead composing machines must contain, in addition to the character codes, all of the necessary functional codes for type font changes, leading, and, of most importance, an approximate line-width count for horizontal justification. This usually requires a complex keyboard device which counts the width of each character, computes the remaining space in the line, the number of interword spaces, and the line deficit. Typical keyboards for linecasting control include the Teletypesetter Standard, Multiface, and Universal; the Monotype; the Roboset; the Linomatic; and the LCC–S Justowriter; in addition to the special keyboards used for photocomposition.

Photocomposing Machines. In the past 15 years, a number of composing machines have been developed that operate on photo-optical principles. Typical of these is the Photon machine invented by Moyroud and Higgonet. (See items 12 and 13.) Figure 32 is a general systems diagram illustrating the operation of the Photon machine. A matrix of 1,440 characters is photographically recorded and stored on a disk which revolves at a fixed rate of either 8 or 10 revolutions per second. A character is identified by a digital code created either by keyboard actuation or recorded on paper tape. This code causes a series of actions to take place. A light source flashes at a precise time and a beam of light is directed through the appropriate character in the matrix and through an optical system comprising a lens turret which will magnify or reduce the size of the character and a right-angle prism which directs the exposed character onto film or photographic sensitive paper. There are several elements common to all photocomposing machines. These are:

1. A matrix of characters in negative form.
2. A light source.
3. A lens or optical system.
4. A magazine or other container for photographic film or paper.
5. A method for identifying the character to be exposed.

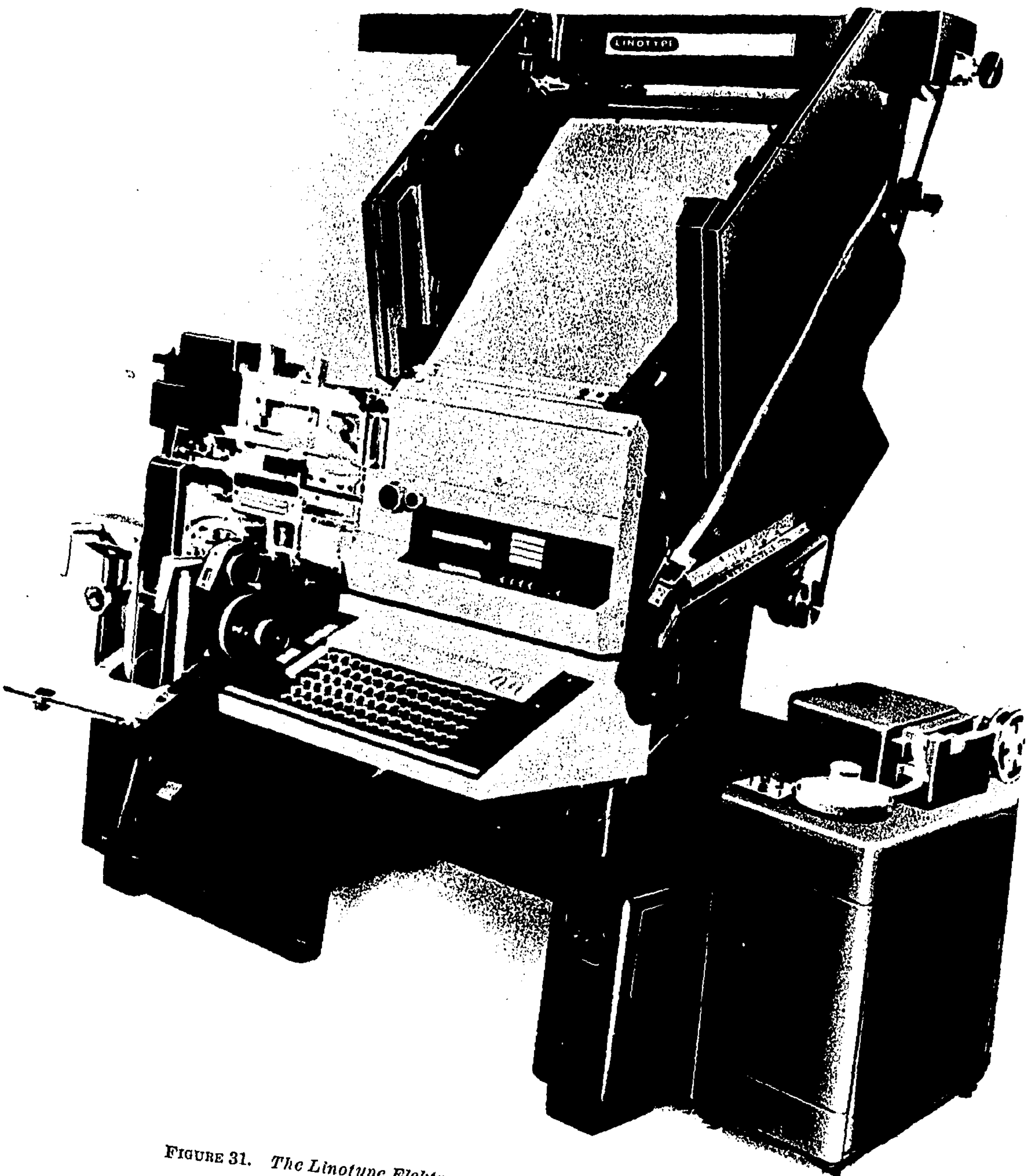


FIGURE 31. *The Linotype Elektron—a high-speed typecasting machine.*

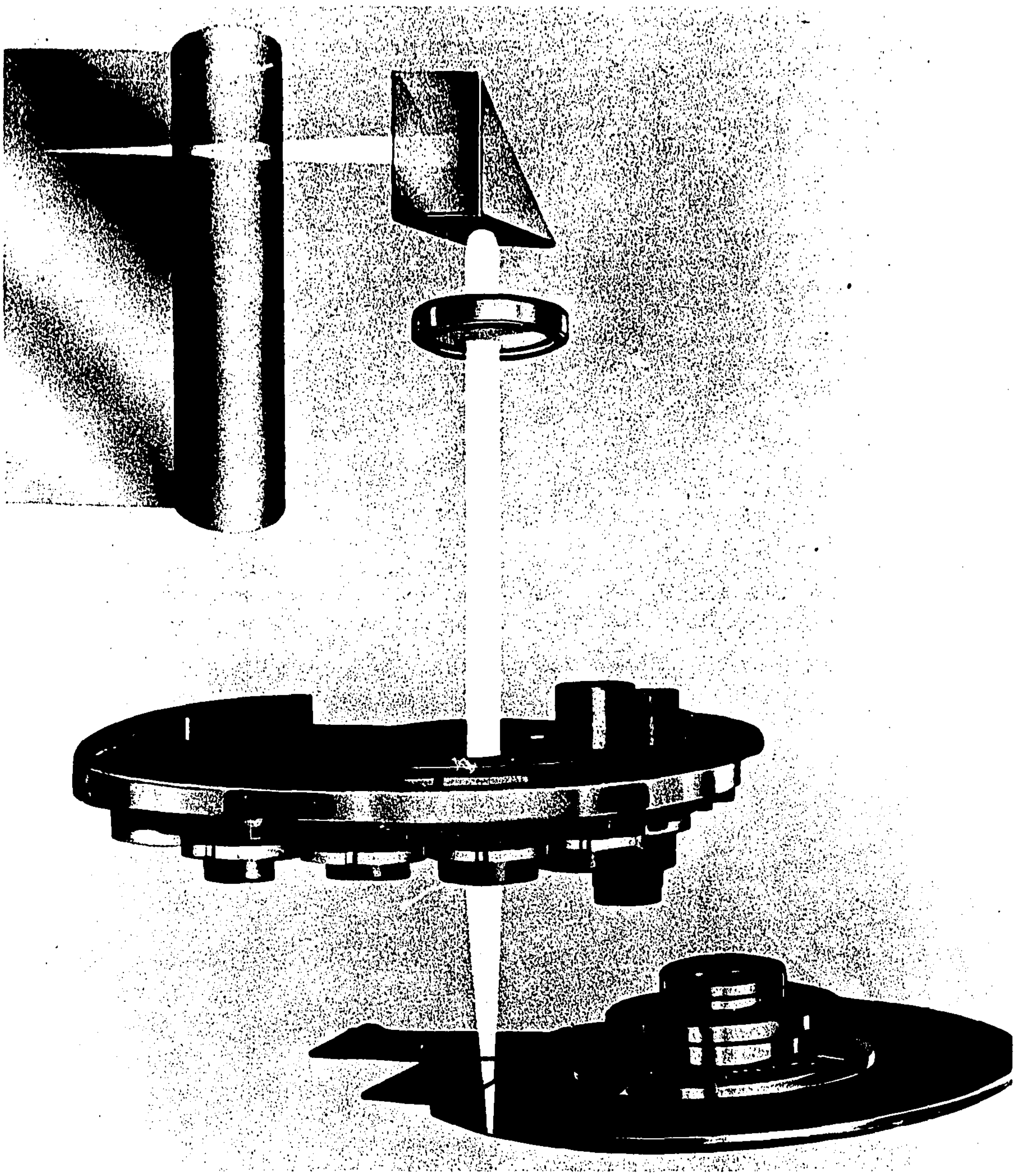


FIGURE 32. *Schematic diagram of Photon type matrix and optical system.*

6. A method for moving the film or paper after each line has been exposed.
7. A method for positioning the character horizontally on the line.
8. A method for quadding, i.e. positioning of the line with respect to the allocated space, e.g. flush left, flush right, centered, or justified.

The Photon can automatically mix from 16 different type fonts of 90 characters each, a total of 1,440 different characters, in each of 12 different sizes ranging from 5 to 95 point. When a character to be photographed is selected, a stroboscopic flash of light (4 millionths of a second) optically scans the corresponding character at the proper instant and exposes its image through appropriate lenses and a prism onto photographic film or paper. One character can be exposed during each revolution of the matrix disk. A variable escapement unit moves a right-angle prism, shown in figure 23, which directs the beam of light to the appropriate position on the film. The mechanical stop and start nature of the variable escapement unit is one of the constraints limiting the speed of the machine. A leading unit operates the film takeup gear drive and provides variable vertical spacing between lines from 0.1 point to 49.9 points for each line.

The Photon 500 series (fig. 33) includes tape-driven photocomposition units which operate from 6-, 7-, or 8-channel tape. The Photon Model 513, which is now being used in combination with the RCA 301 computer for newspaper publication by Perry Publications in Florida, requires only a single code for each character, plus codes indicating line deficit and number of word spaces for use in justification. The Photon Model 560 requires two 8-channel codes for each character, one of which includes the precise escapement required for each character. The 560 has eliminated most of the width count circuitry included in earlier models, relying on the computer to determine the exact escapement values required. (See item 4.)

The Linofilm photocomposition unit (fig. 34) utilizes a 15-channel tape which carries the character code, functional code, and width information. There have been several reported experiments utilizing a Linofilm photocomposition unit

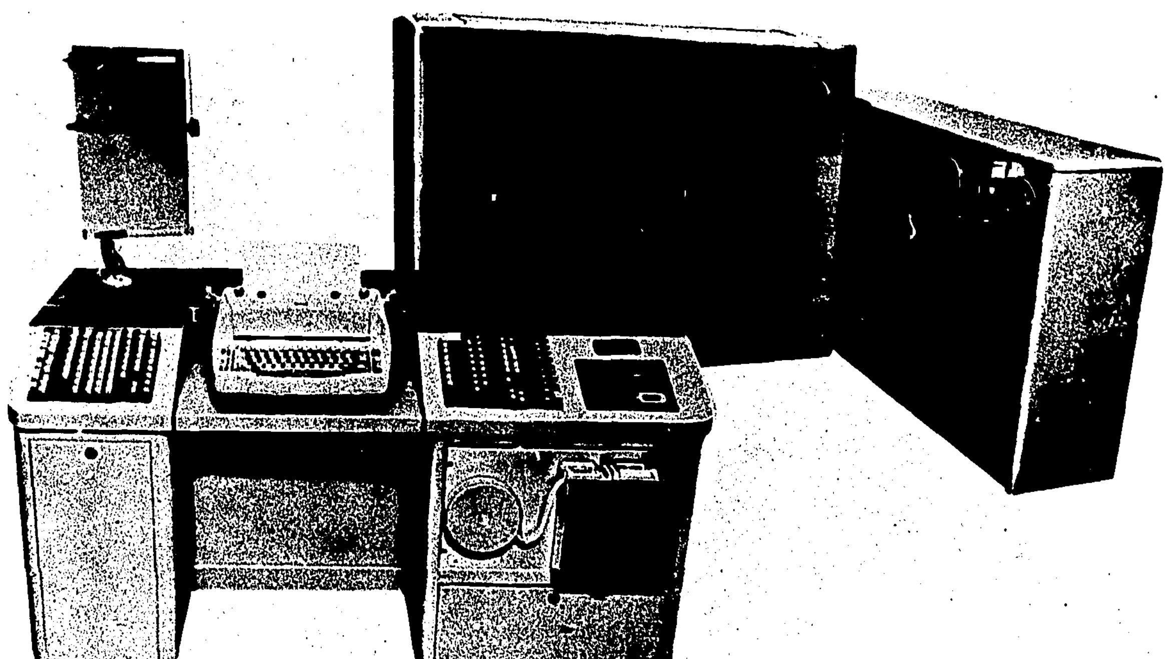


FIGURE 33. *The Photon series 500 composing machine.*

in conjunction with a computer. One of these involved its use in connection with an experimental machine translation project. (See item 18.) Recently, a book on transition probability tables, which were calculated by computer, was composed on a Linofilm photocomposition unit which was cable connected to a general purpose computer. (See item 27.) Because of the obvious inefficiency of operating a 10 to 12 character per second composing machine "on-line" with a high-speed digital computer, a magnetic-tape-to-paper-tape converter was developed which will produce a 15-channel paper tape in the format required to operate a Linofilm photo unit. The approximate cost of this converter is $60,000.

In the Linofilm system, the grid font is stationary at the time the character is being photographed. A light source exposes the entire 88-character font and a shutter system, comprising a series of 8 shutters, masks out all but the one character called for; a series of 88 "lenslets" carry the light to a collimator which places the character in the proper geometric plane. At this point, the image is magnified by a pair of lenses mounted on a sliding bar which provides variable magnification (different point sizes). From here the image is redirected by a front surfaced mirror to the film plane. The mirror is mounted on a sliding bar which moves across the page as the line is being exposed.

Under tape control there are 18 grid fonts, of 88 characters each, mounted on a grid turret (fig. 35). Usually there are 3 similar fonts in different point sizes for each style because the magnification system does not cover the entire point size range from 5 to 36 points. Consequently, there are actually 6 type fonts available over the complete point size range or 18 type fonts available in 4 to 6 different point sizes.

There are a number of less sophisticated and less expensive photocomposing machines presently on the market or about to come on the market. Included in these are the ATF Typesetter which operates at from 5 to 6 characters per second and has a character set of 168 characters; the Monophoto which operates at about 5 characters per second and has a 255-character set; the Alphatype which operates at about 9 characters per second with a 168-character set; and the Intertype Fotosetter which composes at 6 to 8 characters per second and has a 480-character set. A new electronic photocomposer by Intertype has recently been unveiled which composes at the rate of 20 cps, has 20 different point sizes, and 480 characters on 2 disks. It is expected that later models will hold 4 disks thereby increasing the character set to 960 characters.

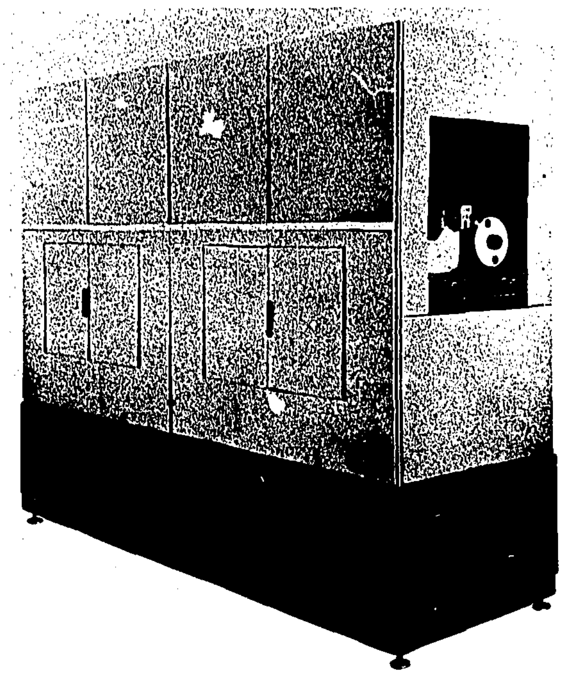


FIGURE 34. *The Linofilm photocomposition unit.*

Sequential Card Cameras.—A technique which has had some application in the library community, especially in the preparation of indexes, utilizes a device commonly referred to as the sequential card camera or step camera. The present (pre-MEDLARS) system for producing *Index Medicus* utilizes the sequential card system. (See item 25.) The indexes to *Nuclear Science Abstracts* and *International Aerospace Abstracts* are produced by this technique as are other indexes, stock lists, directories, and the like. Briefly described, the method involves typing one, two, or three lines of copy (depending on the machine used) in a designated position on a tabulating or EAM card. The cards may be keypunched for automatic sorting by means of electric accounting machine equip-

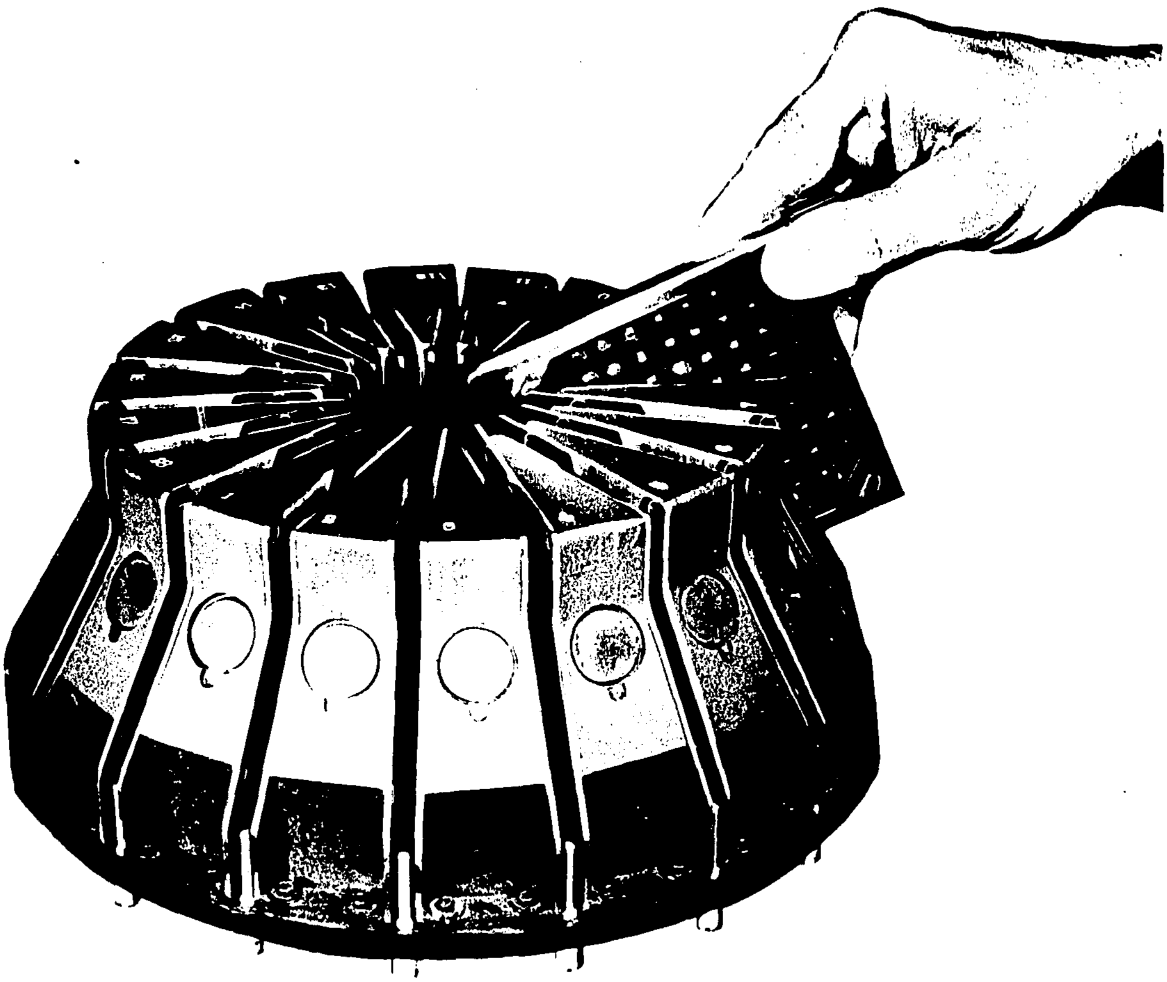


FIGURE 35. *Linofilm grid font turret.*

ment or they may be hand filed for manual updating, additions, or deletions. At publication time, the cards are taken to a device such as a Pitney-Bowes Tickometer to count the number of lines in a given column and to separate the deck into column packages; the column subdeck is then run through the sequential card camera. Figure 36 illustrates the cards with a single line of typing on each and the resulting camera negative output of the sequential card camera. The utility of this system for handling subject heading lists and indexes and even class schedules is quite attractive, assuming that there is no other justification for storing the data in a machine-readable form. The system may even be utilized in connection with tape-operated keyboard equipment for input to a remote computer system. In the case of the present *Index Medicus* system, the initial keyboarding is done on a tape-operated Justowriter recorder and multiple entries are created by reinserting the paper-tape product into the tape reader of the Justowriter reproducer to produce the desired number of cards for sequential card operation. The main difficulty of the sequential card system in this operation is its inflexibility as a tool for searching or preparing special bibliographies. The actual text printed on the EAM card is not punched on the card, and the use of text on the card eliminates some of the columns for punching of data.

OLD PART NO.	NEW PART NO.	DESCRIPTION	PRICE
HL–2420	**50–2420–0**	Knife Feed Bkt. & Bearing	10.50
HL–2430	**50–2430–0**	Selector Cam & Shaft	2.80
	50–2431–1	Clutch Selector Cam	1.15

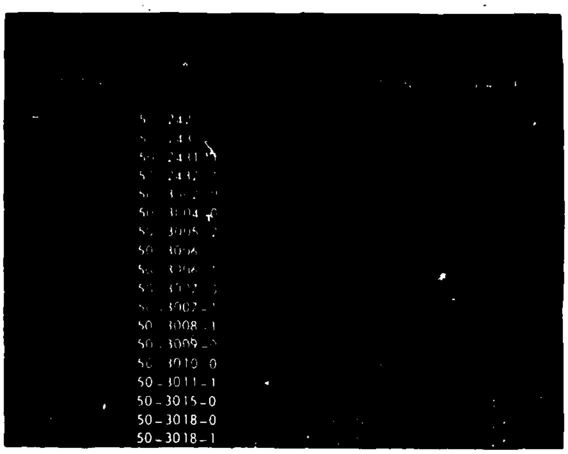


FIGURE 36. ***Output from the sequential card camera.***

High-Speed Computer Output Printers.—Mechanical Printers. High speed has been the principal design objectives for computer output printers. This stems from the fact that the internal processing speeds of general purpose computers are so great. Speeds are being measured today in terms of nanoseconds (billionths of a second) and microseconds (millionths of a second) instead of milliseconds (thousandths of a second). Keeping in step with the high internal operating speed of the computer, mechanical output printers have increased in speed to the point where they are now commonly producing 600 to 1,000 lines per minutes (2,000 cps) depending upon many factors, including the size of the repertoire or character set. These speeds, while still relatively slow as compared to the internal processing speeds of the computer, completely overshadow the present operating speeds of tape-operated type composing machines which compose at the rate of from 5 to 20 characters per second. A one-thousand-line per minute printer with 120 print positions can create text at the rate of 2,000 characters per second. The character set used in this extremely rapid process is, however, very small and it is manipulated without any of the control devices (justification, leading, etc.) used by typesetting machines. The result is a product of low typographic quality. It may be possible to improve the typographic quality of the mechanical printer and achieve a compromise between speed and quality. An example of this has already been accomplished by modifying an IBM 1403 Printer and increasing its character set to 120 characters. (See item 8.) This modification of the print chain has reduced the normal printing speed by about 60 percent. Figure 16 includes an example of the quality which can be produced from such a modified chain printer. Figure 37 is a diagrammatic view of an IBM Printer.

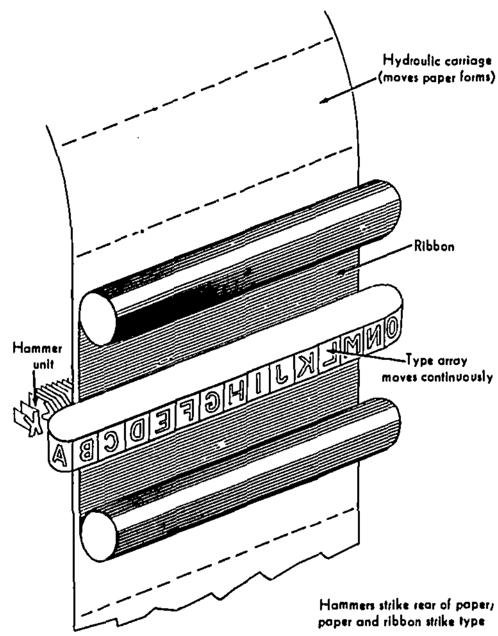


FIGURE 37. ***IBM 1403 Printer—diagrammatic view.***

A number of articles have been published describing the evolutionary development of the mechanical printer from the single typebar to the type wheel, the type roll, and the type chain of the IBM 1403. Items 7, 15, 32, and 40 are representative of this literature.

A further development of high-speed computer printers is the matrix printer. This device forms characters on paper by a pattern of dots produced by a matrix of needles or styli. A typical system employs a 5 by 7 inch matrix for a dot pattern having a maximum of 35 dots. One of the basic difficulties is that every character in this system requires 35 bits of information. The product of this system is far from graphic arts quality. Developments in this field are widely known.

Electronic and Optical Composers. The basic speed limitations of all present-day graphic arts quality composing machines are mechanical in nature. In Linofilm, the shutter, magnification system, and escapement system all have mechanically moving parts with considerable mass to overcome. In Photon, the matrix disk revolves and, on a start-and-stop basis, the variable escapement unit moves an amount determined by the character-width control circuitry. The same is true of other photocomposing machines. As far as the hot-lead machines are concerned, the mechanical problems of the delivery and return of the brass matrices, plus the "recording" time of pouring the molten lead over the assembled matrices and ejecting the finish lead slug, are speed-limiting.

Electronic and optical composing techniques have been developed which are capable of operating at high speed by overcoming mechanical limitations. Analysis shows that the basic problems common to output printing are (1) forming the image, (2) locating the image, and (3) recording the image.

The composing methods described below are methods of forming the output printing image by electronic and optical techniques. Essentially, there are five distinct electronic and optical methods of forming or generating characters from digital codes.

1. The character can be formed by passing an electron beam through a stencil-like cutout in the shape of a character located between an electron gun and the face of a cathode ray tube (CRT).
2. The monoscope method of character generation utilizes an electron beam that hits a metalized target within a CRT, with characters printed thereon, causing a video signal corresponding to the desired character, which signal is amplified and displayed on a separate CRT face.
3. A digitized matrix is generated wherein the character selected represents a series of intersection points with the raster lines of the matrix displayed on the face of a CRT.
4. A CRT scans a character mask, which is larger than the face of the scanning CRT and outside of its envelope, by using an optical tunnel, and the resulting video signal is displayed on the face of a separate CRT.
5. The characters are generated optically by flashing a light behind a rectangular matrix and directing it onto film by means of a pair of parallel mirrors, a traveling lens, and appropriate electronic timing circuitry.

Examples of both photo-optical and electronic character generation are briefly mentioned here to illustrate these methods.

1. *Photo-optical Character Generation.*

In the GRACE [31] system, being developed for the MEDLARS program of the National Library of Medicine, and Photon Corporation's commercial version known as ZIP, mechanical movement and mass have been reduced to a minimum in order to increase speed to a maximum. (See item 28.) The principal moving parts in ZIP are (1) a traveling lens that traverses the page horizontally composing a line with each sweep and (2) the film advance mechanism. The mass of the lens has intentionally been kept small. The matrix plates containing the character images are stationary. By means of an optical device comprising two parallel mirrors, the characters in a vertical column are directed to a single horizontal base line. (See fig. 38.) The GRACE and ZIP systems (see fig. 39)

[31] GRaphic Arts Composing Equipment.

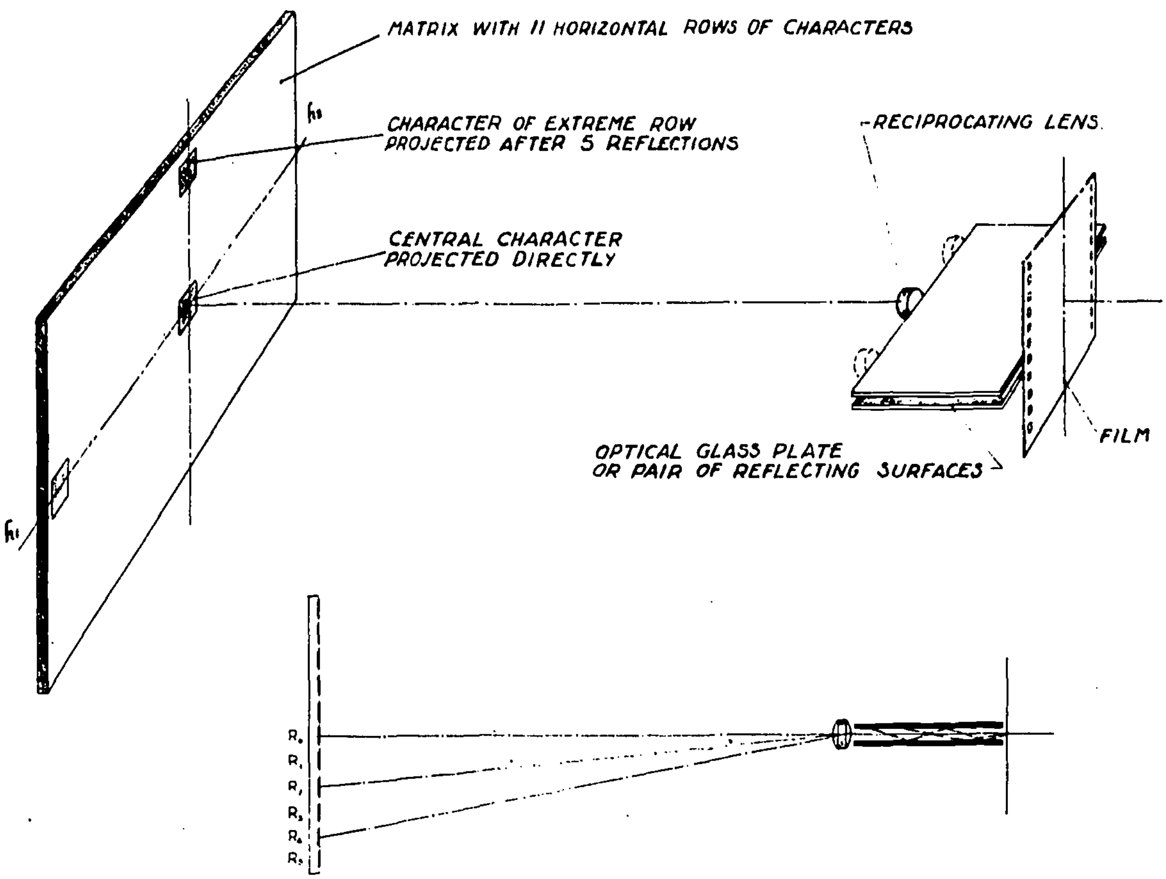


FIGURE 38. ZIP *system diagram showing parallel mirror surfaces.*

utilize a separate flash tube located behind each character. The flash tubes will discharge at a precise time in accordance with the electronic circuitry when the traveling lens (see fig. 40) is in the proper position with respect to the character matrix and optical device. The electronic timing system takes into account the proper escapement for each character, the value of word spaces, the location of each character on the matrix plate, and the position of the traveling lens.

2. *Electronic Character Generation.*

Various electronic techniques for generating characters have been in use for several years. The commercial applications of these techniques have not been in publications requiring graphic arts quality.

An example of a successful commercial application of computer output recording by electronic character generation is found in the SC 4020 high-speed microfilm recorder made by General Dynamics/Electronics. (See items 11 and 25.) The SC 4020 displays alphanumeric data on the screen of a special CRT called the Charactron-shaped beam tube. To record the material presented, the data on the tube face are projected through an optical system to a high-speed 35 mm camera. The unit could also simultaneously record the image on

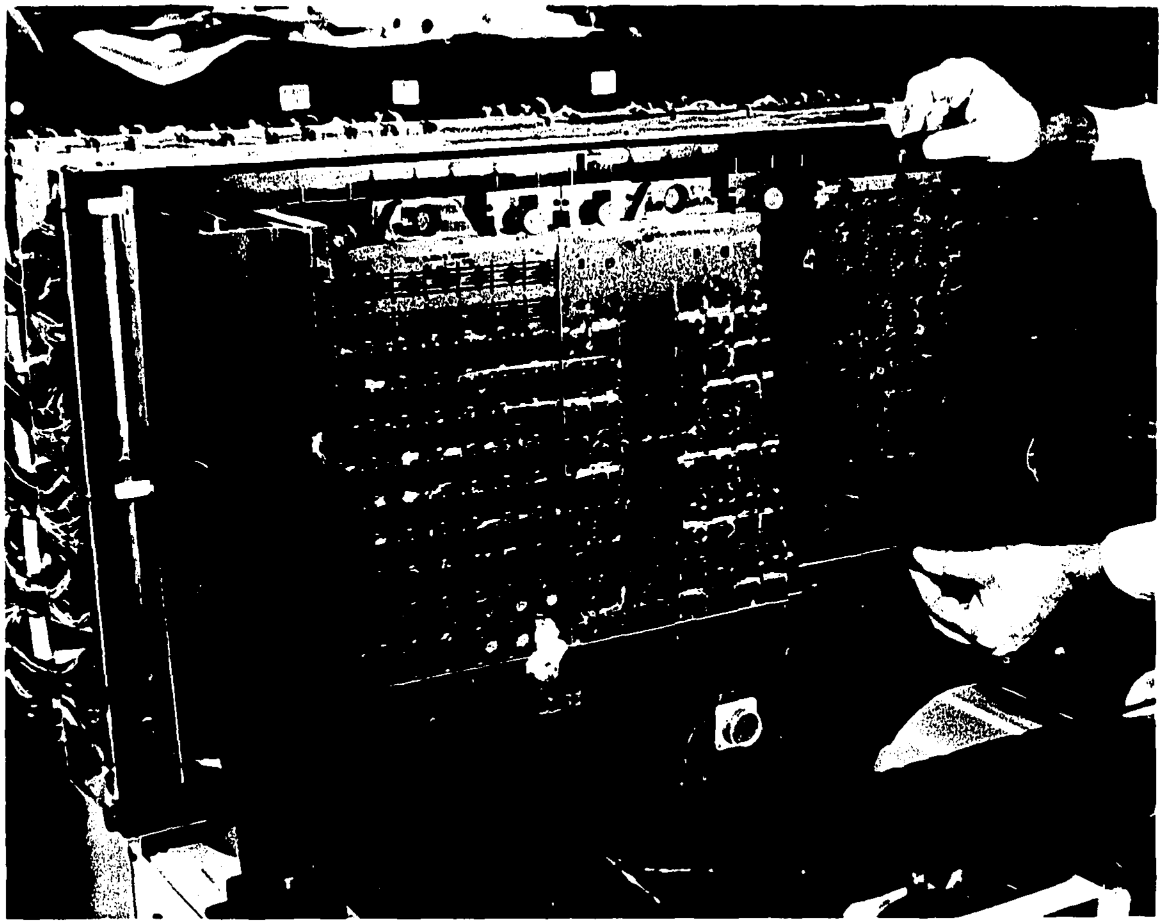


FIGURE 39. *Character matrix for Photon* GRACE *and* ZIP.

9½-inch wide photorecording paper with an optional unit. It composes characters and symbols at a rate of 17,400 per second. The characters are formed by directing a beam from the electron gun at a thin metal disk which may have as many as 64 different characters arranged in an 8 by 8 inch matrix which is cut out like a stencil. Selection plates, located between the electron gun and the matrix, are supplied with d.c. control voltages which direct the beam at the desired character; then horizontal and vertical deflection circuits deflect the selected character to the appropriate spot on the CRT display. This technique has also been successfully combined with a dry process Xerographic printer which has achieved a composing rate of 1 million characters per minute. Other examples of nonmechanical computer output recording devices employing character generation are described in the literature, e.g. items 7, 24, and 32.

A number of proposals have been made by various companies suggesting that a graphic arts quality composing machine employing various electronic character generation techniques could be developed. Firms known to be working in this area include A. B. Dick Co., CBS Laboratories, Mergenthaler Linotype Co., and Radio Corporation of America.

The greatest problem to be overcome in developing an electronic graphic arts quality computer output printer is that of positioning the images both horizontally and vertically on the page within the close tolerances required. An additional problem will be the provision of multiple type fonts with intermixing capability within the boundary of reasonable cost limitations.

The traditional method of recording the image has been by the mechanical impact of the hammers and typebars against an inked ribbon and onto the paper. A variety of new recording techniques are in use and in various stages of development which will remove the mechanical limitations on speed. These developments include electrostatic printing, magnetic printing, smoke printing, thermal recording, thermoplastic recording, and photographic recording. Many of these techniques can be applied to the making of replica copies of existing documents as well as to generating the initial copy of a document from an electronic or digital store. Such techniques for graphic storage and replication of copies are the subject of Dr. Alexander's paper and will not be discussed here. Based on the system requirements for edition printing of varied library publications, the method of photographic recording onto film or photosensitive paper is an adequate recording technique for preparing plate-ready copy for a sizable printing run.

Programming and Systems Considerations

The complexity and variety of equipments and processes described above require that, in their use, much care and attention be devoted to factors of intermachine relationships and to programming operations. An example of the kinds of problems encountered in making these machines work in an operating situation is given in this section.

Input Preparation.—The problem of preparing input for a mechanized library which contemplates output printing for publication differs from that of other data processing systems primarily in the increase in the number of characters

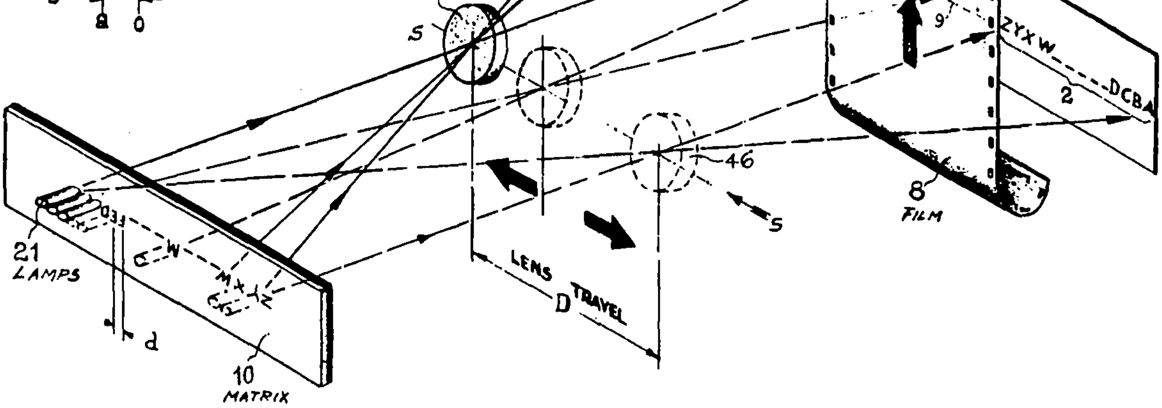


FIGURE 40. ZIP *system diagram showing traveling lens.*

735-898 O—64——13

to be dealt with and in the inclusion of control functions which have no special meaning to the computer itself. It is likely, however, that the actual input preparation will be kept as simple as possible, leaving tasks of horizontal justification, hyphenation, reformatting for publication, insertion of typographic control functions, updating, and sequencing for publication to the computer. Consequently, the major considerations relating to input preparation will be reduced to matters concerning the form of the input, i.e. whether it appears on punched cards, punched paper tape, or magnetic tape and the codes and coding methods used. Hopefully, these can be standardized, but it is likely we will have to cope with a wide variety of inputs and formats for some time to come.

Input Code Conversion.—If the machine-readable record of the input preparation device is not directly readable by the computer, the intermediate step of off-line conversion may be required. It may be necessary, for example, to convert the paper-tape or punched-card output of a keyboard device to the specific paper-tape or punched-card code used in a given computer. Alternatively, it may be necessary to convert paper tape to punched cards or to magnetic tape, or conversely, punched cards into paper or magnetic tape.

Fortunately there are a number of converters already available for this purpose. Examples include the General Instruments C–750/026 tape-to-card converter; Friden tape-to-tape converter; IBM 046, 047 tape-to-card converter; IBM 063 card-to-tape converter; Systematics tape-to-card, card-to-tape, or tape-to-tape converter; Addressograph-Multigraph 941 card-to-magnetic-tape converter; IBM 7765 paper-tape-to-magnetic-tape converter; Linofilm magnetic-tape-to-15-channel-paper-tape converter; Digitronics D300 magnetic-tape-to-paper-tape converter; and the Electronic Engineering Co. of California magnetic-to-paper tape, paper-tape-to-magnetic-tape, and paper-tape-to-paper-tape converters. The use of code converters, however, requires very careful attention to code combinations and to the reservation of codes for control functions and format and geometry of the paper or magnetic tapes.

Input Processing.—A complex computer system will serve many purposes in the mechanized library other than typographic composition. The majority of input processing problems are therefore the concern of those responsible for file conversion. Typographic composition will however present problems of encoding to handle lexical information containing character sets in multiple fonts. The internal character code may use 8 bits with as many as 256 code combinations (plus a 9th bit used as a check bit), a 7-bit code having 128 code combinations (plus a check bit), or a 6-bit code, having only 64 code combinations plus a series of mode changes or so-called precedence codes. The final selection of input codes involves many complex factors including computer word length, file organization, and ease of manipulability. Of no small importance is the objective of having a standard single code configuration. Standardization offers the advantage of compatibility at the cost, for some applications, of increasing complexity. Conversion of the input code to the internal code for processing purposes is essentially a simple table lookup problem.

Alphabetizing.—When processing text requiring character sets beyond those of the standard 47-character set of the IBM 024 or 026 keypunch, the difficulties of alphabetizing are greatly increased. A simple example is the alphabetizing of proper names containing abbreviations such as St. Andre or St. Claire. Such names are normally alphabetized as if the word "Saint" were fully spelled out and correct filing therefore involves addition to the actual text. It is also necessary to eliminate from the text functional codes which have no bearing on the alphabetizing sequence. If a discretionary hyphen is utilized in the keyboarding, it should be given no value in alphabetizing. Most of the problems of alphabetizing can be solved by internal code conversion. The process of converting one character to another (e.g. capital A to small a) would essentially be a table lookup operation, even where precedence codes are employed. The problem of converting "St." to "Saint" is more difficult since it depends on awareness that the information field might contain proper names which start with an abbreviation. The program must test for the presence of abbreviations, for example, by looking for a capital "s" followed by a small "t" and a period. The method utilized for alphabetizing will have a direct effect on the interface between output printing requirements and the console display. The reason for this is that some of the possible alphabetizing subroutines are nonreversible. Since the problem of alphabetizing is inherent in library file systems and falls within the province of other papers on file conversion and file organization, we will not attempt to go into further detail in this paper.

Automatic Hyphenation and Justification.— It is quite likely that the precise format of the output publications in terms of how the information is to be arranged, which column width is to be used, and which elements of information such as author, title, date, are to be included will not be determined by the input keyboarding. For this reason, it will be necessary for the computer to instruct the output device concerning the width and length of columns, insertion of illustrations, page numbers, etc. More important and more complex, however, is the process of determining how much information will be contained on a single column line, if this is not determined in the input keyboarding. Since each character in graphic arts typography has a distinct character-width value for the point size to be set, the computer must look up this value as each character is set and cumulate the total to determine how many words can be composed per given column line. It would be convenient if the number of words always came out even; however, this is simply not the case. Lack of correlation between character count and line length complicates the problem of hyphenation. A solution of this problem might be to begin the word requiring hyphenation on a new line and leave the right-hand margin of the preceding line unjustified. This probably will not be an acceptable solution, however, because many words are so long that the preceding line might not only be unjustified but might even be almost or completely blank. Obviously, as the column width is made smaller the problems of hyphenation and horizontal justification become more acute.

In the past several months, a number of important breakthroughs have been made in computer techniques for achieving justification and automatic hyphenation. For example, different techniques are in use by Perry Publications, the *Los Angeles Times*, and the Oklahoma Publishing Co. (For additional information, see items 14, 30, and 41.)

The Perry Publications System—A Dictionary Lookup Approach. This system includes an RCA 301 computer with 20,000 internal memory units, a paper-tape reader operating at 1,000 characters per second, a paper-tape punch operating at the rate of 100 characters per second, and 6 magnetic tape drives. The equipment configuration also includes a 1,000 line per minute mechanical printer for performing functions other than hyphenation and justification. The computer operates in the simultaneous mode, that is, it can be processing for hyphenation or justification at the same time that it is reading or punching tape. Since the speed of the paper-tape punch would otherwise be a limiting constraint, the processing time required in this approach to hyphenation can be overlapped to capitalize on this limitation. The system utilizes a dictionary lookup approach wherein approximately 30,000 to 40,000 words are stored on magnetic tape on 4 separate tape drives. Obviously, only the most frequently occurring words are stored. For example, a dictionary containing 13,000 words would include hyphenations for 90 percent of all words having over 5 letters based on a 2-week sample. Since the RCA 301 computer can search tape in both directions, the tapes are always maintained at a given "homing" position and the most frequently occurring words, based on their first letter, are located closest to this homing position on each of the 4 tape drives. An index to the positions of the dictionary on the 4 magnetic tape units is maintained in core memory using the first 2 letters of the word as a key.

If words can be correctly hyphenated by dividing after the third, fifth, seventh, or ninth letter, they are not included in the dictionary since it is reported that 48 percent of the commonly used words are correctly hyphenated following this 3–5–7 rule, and that 90 percent of the hyphenations made following this rule were incorrect by only one character. The computer proceeds through the following steps until line justification is obtained:

1. Justification is attempted by expanding space bands. (Hot metal system)
2. Hyphenation is attempted at key prefixes and suffixes determined by the lookup within the computer (e.g. sub-, pre-, -tion).

3. Hyphenation is attempted by looking for the word in the stored dictionary described above.
4. Justification is attempted by adding thin spaces between each word.
5. Hyphenation is completed by arbitrarily dividing the word after the third, fifth, seventh, or ninth letter if the word is not in the dictionary.

The Los Angeles Times—A Logic Approach. The *Los Angeles Times* system utilizes an RCA 301 computer with 20,000 characters of core memory, an RCA 1,000 character per second paper-tape reader, a special Soroban paper-tape punch which operates at 300 characters per second, and an RCA paper-tape punch operating at 100 characters per second. The additional high-speed punch is required to overcome peakload conditions imposed by the size of the Sunday edition of the newspaper. One of the interesting aspects of the *Los Angeles Times* system is the fact that it does not require any external memory, such as magnetic disks, drums, or tape drives. The output printers are hot-metal linecasting machines which operate from 7-level paper-tape readers. The system generally involves one hyphenation per 7 lines of text. Speed is approximately the 300 character limit imposed by the punch.

The hyphenation system does not rely on the dictionary approach but rather on logic tables, which occupy only 5,000 core memory positions, and which handle all words, including proper names. Although this system does not always divide words according to Webster, test runs indicated that over 99 percent of the hyphenations were reported acceptable following the rules of word divisions set forth in the introduction to Webster's unabridged dictionary. This percentage was calculated by dividing the total lines correctly hyphenated by the total number of lines hyphenated. The logic is based on the following principles. First, vowel and consonant patterns in a word are classified into one of four basic types. The computer then scans key letter sequences to see if they follow the rules governing the type. If so, an immediate solution is reached. For example, prefixes and suffixes which are commonly used can automatically determine hyphenation. Where exceptions are indicated, they are defined and analyzed by following special logic subroutines such as testing against letter sequence tables. In this way, the nature of the exception is defined and a solution is reached. Various techniques are employed in using letter sequences as a key to phonics. Among them are table lookups in which the cumulative effect of any three letters of the alphabet can be weighed and various paths taken as a result. Comparisons are also used in sensing ahead for other vowels and in determining the beginning or end of a word.

In dealing with the exceptions, which are perhaps more numerous in English than in any other language, the computer sometimes leaves the three letters it is directly concerned with and backs up or jumps forward two or three letters as a means of making its analysis as inclusive as possible.

The Oklahoma Publishing Co.—A Table of Probabilities System. This system utilizes two IBM 1620 computers with 20,000 characters of core memory, IBM tape readers operating at 500 characters per second, and tape punches operating at 50 characters per second. (The latter are to be replaced with 150-character-per-second punches.) The system is being operated on an experimental basis.

The hyphenation program begins with an edit of the word to determine the number of syllables, and in some cases determines the hyphen point, or the inability to hyphenate at a point. It next determines the probability of hyphens occurring between any two letters in the word and hyphenates at the most probable point. A limited number of abbreviated words are stored in a table for lookup. Accuracy of 94 percent is reported for this program based on current production tests.

The B.B.R. System. Several years ago, an automatic hyphenation system was developed in France by Bafour, Blanchard, and Raymond. (See item 2.) They point out that the empirical rules for hyphenation in the French language have been established and effectively proven. Examples are as follows:

- Not to cut after less than two letters.
- Not to cut so as to leave less than three letters.
- Not to cut after a consonant followed by a vowel.
- Not to separate two vowels.
- Not to separate certain vowel couples forming an inseparable doublet.

Not to cut after a vowel if it is followed by two consonants which form an inseparable doublet.

Not to cut before the letter *y*.

Not to cut before a punctuation sign.

Not to separate two numbers.

It would seem that the system being utilized by the *Los Angeles Times* utilizes some of the same principles.

The Discretionary Hyphen System. A manual technique for hyphenation, either with or without a computer system, known as the discretionary hyphen system was first tried by Louis Moyroud, one of the coinventors of the Photon machine. This technique requires the keyboard operator to insert a discretionary hyphen in every long word. The computer or output printer can then utilize or disregard the discretionary hyphen during the actual process of completing horizontal justification. Although the use of discretionary hyphens is estimated to increase keyboard operation time by 2 to 5 percent and storage requirements by 2 percent, the simplicity of the technique is somewhat attractive, especially for noncomputer output printing systems.

It must be emphasized that there is no error-free technique for hyphenating a word, either automatically or manually. Human keyboard operators do not have perfect recall of the content of Webster nor do they always refer to a dictionary when in doubt. A certain degree of error is tolerable, in any event, as there are a variety of ways in which corrections can be made.

A semiautomatic method for line justification and hyphenation has been developed by the Compugraphic Corp. and is incorporated in a special-purpose computer-like device known as Linasec, which sells for approximately $27,000. Linasec reads unjustified paper tapes produced on simple monospacing paper-tape keyboards and automatically justifies each line unless it cannot be justified without dividing a word. At this point the line is displayed and the machine stops for human intervention. Since this is only a $27,000 device and not a $200,000 to $2 million general purpose computer, it is practical to allow the machine to interrupt and wait for an operator to decide where the hyphen should be placed. Obviously, such a man-machine interrupt feature would not be practical on, for example, an IBM 7090 computer system which costs from $400 to $600 per hour to operate.

Output Formatting.—When the computer has to format data in columns, the number of lines which will fit in a column can be computed. If a 3-column format is to be utilized, it is possible to rearrange this information on the output tape so that the composing machine can compose three columns across simultaneously and thereby eliminate the need for manually stripping up three separate columns of paper or film. It is understood that this technique will be utilized in the MEDLARS system. The computer can assign page numbers, allow predesignated spaces for inserting drawings, photographs or other graphic information, and can insert subject headings and column headings automatically. By minute variation in the leading between lines, the computer can easily accomplish vertical justification so that the bottom and top lines of all columns on the same page will be flush.

Typographic Functional Codes.—A predetermined set of rules can be programmed into the computer to instruct the output printing device when to change fonts and point size, how much lead to leave between lines, how to determine column width and the like. In the printing of bibliographic information this can usually be tied to specific elements of information such as the author, title, call number, corporate author, which can be composed in bold face, italics, all uppercase, underscored, and so on. The functions for quadding can also be part of the program. Alternatively, the input tapes may contain all of the necessary typographical control functions stored either together with or separate from the data itself within the computer external memory systems until the point in time that the specific output is required. This will undoubtedly be necessary in cases such as type font changes to indicate special symbols, Greek and Cyrillic alphabets, superscripts and subscripts, etc. Simple format instructions such as boldface for title or italics for name of journal can be provided automatically.

Output Code Conversion.—As mentioned earlier, the code produced by the input keyboard may differ from that which can be read by the computer, from the internal code which the computer will use in its processing, and will probably also vary from the code required by the output printing device. As indicated in the previous section on "Output Printing Equipment," existing equipment operates from a wide variety of tape inputs, both paper and magnetic. The paper tapes include 6-channel TTS, 7-channel Justowriter, 8-channel IBM Flexowriter, 8-channel Photon code, 15-channel Linofilm, and 31-channel Monotype and Monophoto codes. This suggests that a considerable amount of code conversion on the output side will have to be done either by the computer or an off-line magnetic-tape-to-paper-tape converter. Even where high-speed magnetic-tape-operated photocomposing machines are utilized, the format of the magnetic-tape store of the computer may differ from that of the output printer. Custom design of an output printer for a particular system can, of course, bring together the appropriate tape transport with special decoding logic at the input side of the printer.

The Integrated Systems Approach

The foregoing discussion suggests that successful automatic output printing for the library, both for computer and noncomputer operations, is in large measure a matter of bringing together a variety of equipments and processes. Because this is so the integration of these systems and processes—the systems approach—is of considerable importance to output printing systems design.

A well-thought-out system design will include careful consideration of a variety of factors and will not be satisfied with simply a piece of terminal equipment for the computer. The printed output of the library represents an important part of its services and functions, and it will become even more important with the mechanization of the internal stores of the library. The variety of printing requirements in the library is a challenge to printing system design, but not an impossible one. Careful choice of equipment and processes and provision for their assemblage in an integrated whole, viewed from the demands of the entire system, can result in the successful solution of the problems raised by these requirements.

The techniques of system design, as elaborated by those most concerned with large-scale problems of this nature in the communications and military weapons industries, usually involve a sequence of steps such as: (1) system planning; (2) development of the system requirements and the system design; (3) implementation of facilities; and (4) pilot plant adjustments and full-scale operational adjustments. Some of these steps will be examined briefly here in relation to library output printing design.

System Planning.—In a very real way this conference on library mechanization is part of the system planning phase. At the outset of the library mechanization project something must be known about the objectives of the system, the functions to be performed, the technical and economic feasibility of providing these functions, and the organization and assignment of responsibilities for getting solutions underway.

Each conference paper has stressed the need for systems planning at the level of the general library community if successful library mechanization is to be widespread. The network concept is not new really, for operational networks exist now in the library community. Too many of the functions carried on in a mechanized library system are affected by what goes on outside its own operations to ignore the broader library environment. It is perhaps unnecessary to point out that the Library of Congress has a major hand in supplying the bulk of bibliographic tools used in the internal operations of all libraries in the United States. With this as a starting point, it appears that there is considerable hope for general library mechanization.

Library output printing is particularly concerned with these environmental factors. It was pointed out above that there is a direct relationship between the original encodement of bibliographical data in the input phase and the product of the output printing process. If this original encodement is to take place in decentralized contributing libraries within a common network, we can readily understand how important matters of process and equipment compatibility can be. The great national libraries have the serious responsibility of designing their own mechanized systems in such a way as to provide leadership for other libraries while at the same time not preventing them from participating in mechanization by choosing, at the national level, equipments or processes with which the smaller libraries could not afford to coordinate. It is encouraging to see, in this regard, that methods of encodement in machine-interpretable (not simply machine readable) format are being developed which adapt traditional cataloging processes. The methods can be applied to the simpler input equipments which might be used at the local level. Coordination of this potential input with the output printing phase of national (and local) library operations is essential.

Specification of Requirements.—System planning must also include a study of operational details of the kind discussed in the sections of this paper which dealt with the products and aspects of library output printing. Such study will lead to detailed specification of requirements for machinery, processes, and manpower, which, by reason of the care taken to relate them to both the environmental and internal system factors, should provide the most economical and reasonable overall design.

The usual procedure is to divide the system into subsystems for this detailed analysis. In the case of library mechanization the output printing subsystem is, as we have indicated, of more than passing importance. It is influenced by and, in turn, influences the input subsystem and is of direct importance to the file manipulation and storage subsystems. Decisions in all these subsystems should not be made without reference to the others. This will assure internal compatibility of all system elements in the mechanized library.

Implementation.—Reconciliation of subsystem requirements on a technical basis will eventually result in one or a series of technically feasible designs. Three other steps must then be taken. Compromises must be made for the sake of economic feasibility; special equipments, if needed, must be designed and produced; and installation of the pilot operations must be carried out. In library mechanization, economics will be of importance in the output printing subsystem, since the cost of this operation will not be negligible and since it is the means through which the mechanized library will seek a broader support in patron-oriented products.

After precise technical specifications are prepared, equipment orders must be placed. There must be allowance for adequate time for adjustments at the factory and, after installation, at the library. Planning for adequate time to set up and check out the equipment of each individual subsystem and of the overall integrated system before production commitments commence is usually overlooked. Experimental checkout cannot be excessively prolonged, yet too early a commitment for meeting production schedules will result in errors, breakdown, and confusion. Some projects involving automated lexical systems have experienced this failure in planning and have been discontinued because of it.

The Systems Team.—The output printing problems involved in library mechanization are an excellent example of the need for the broadest possible approach. They range from problems of source data encodement to such intricate details as computer word length. Their solution, therefore, cannot be sought in any narrow specialism but must be the result of many points of view. Teamwork in framing the problems of library mechanization and in seeking their solution is absolutely necessary. The systems team must be composed not only of high-level technical and management personnel, but it must also include a broad sampling of talent from a multiplicity of disciplines and specialties with an especially generous proportion of librarians.

The most difficult task facing the library administrators who will be responsible for formulating and implementing an automation program will be the selection and organization of the systems team. In most complicated design problems the limitations are not usually in the lack of equipment, but in the knowledge of how to put a complicated array of techniques, equipment, and manpower to the most effective use.

Conclusions

Output printing is an important aspect of the mechanization of a library. Indeed, it may very well be the key factor in the determination of the economic feasibility of library automation. In this report we have attempted to describe the users' needs for the potential output printing products from an automated library store. We believe that these needs are quite realistic and important and, in some cases, are not currently being satisfied. Output printing permits the content of the mechanized library store to be communicated to a multiplicity of users. Although publications are now being produced by conventional techniques, the problems of updating and cumulation have made many of them too expensive to produce as frequently as required. We have examined the relationship between output printing and the so-called console display and have concluded that the requirements are quite different. In most cases, output printing involves multiple type fonts and high-quality typography. These requirements place special constraints on the system; however, they do not eliminate the possible need for console display for man-machine communication.

The relationship between the output printing subsystems and other operations of the library clearly indicates the necessity of coordinating the design of these operations in library mechanization. This is especially true of the input systems and formats. The development of means for encoding bibliographical data in order to provide records which are machine interpretable as well as machine readable is of prime importance. It is equally important to accomplish this through the use of machines which are reasonably inexpensive and available to smaller libraries.

There are a number of automatically operated composing machines capable of handling applications with a wide range of complexity and composition volume. With respect to computer software and other system requirements, including the interface between various elements, no major breakthroughs are required. The state of the art in output printing today is not the limiting factor.

A considerable amount of pioneering work has already been done in the field of output printing from a mechanized store. Of most significance for the library community is the MEDLARS project of the National Library of Medicine. This Library is to be commended for its foresight in viewing the problem as basically a systems problem having many aspects. They have also made an important contribution in identifying a void in available high-speed, graphic arts quality composers and in filling this void by sponsoring the development of such a machine (GRACE). The entire publication industry will eventually benefit from this work.

The newspaper industry has also pioneered in the use of computers to solve some of its typesetting problems. With the computer performing the functions of justification and automatic hyphenation, the initial keyboarding operations have been simplified and speeded up. The newspaper industry has found it practical to utilize computers because of their inherent high speed in spite of the fact that there is still no commercially available high-speed, graphic arts quality composer. These computers are generally used to prepare punched paper tapes to operate banks of relatively low-speed, hot-lead linecasting machines. In some cases they are also driving low-speed photocomposing machines with computer-produced paper tapes. Manufacturers of computers, peripheral devices, and composing machines have spent large amounts of their own capital to explore markets and to develop high-speed typesetting equipment.

To exploit output printing as a tool of the library of the future there are foreseeable problems in organizing a plan of action, defining the precise system requirements, and marshalling the necessary resources. In addition to financial support these resources should include experts in library technology, systems engineering, computer systems and programming, operations research, and output printing.

Bibliography

1. Ashworth, B. P. Refresher course: the historical background to line casting; construction of machines for line casting. British printer, v. 74, Sept. 1961: 132, 134, 136; Oct. 1961: 132, 134, 136.
2. Bafour, George P. A new method for text composition: the B. B. R. system. *In* Association of Printing Technologists. Printing technology; proceedings of the conference, v. 5, Nov. 1961: 65–76.
3. Bagg, Thomas C., *and* Mary E. Stevens. Information selection systems retrieving replica copies; a state-of-the-art report. [Washington, U.S. Dept. of Commerce, National Bureau of Standards] 1961. 172 p. ([U.S.] National Bureau of Standards. Technical note 157)
4. Barnett, Michael P., K. L. Kelley, *and* M. J. Bailey. Computer generation of photocomposing control tapes. pt. 1. Preparation of Flexowriter source material. American documentation, v. 13, Jan. 1962: 58–65.
5. Berul, Lawrence H. Automated composing systems and techniques; a state of the art report for the National Bureau of Standards. Wakefield, Mass., Information Dynamics Corp., 1963. 1 v. (various pagings)
   Unpublished report.
6. Berul, Lawrence H. Selecting a system for producing higher quality announcement journals. Wakefield, Mass., Information Dynamics Corp., 1962. 53 p.
7. Davie, W. A. J. High-speed printers. *In* British Institution of Radio Engineers, *London*. Journal, v. 20, Sept. 1960: 675–683.
8. Dyson, Malcolm G., *and* Michael F. Lynch. A computer-produced express digest. Columbus, Ohio, Research and Development Division, Chemical Abstracts Service, Ohio State University [n. d.] 28 p.
   Processed.
9. Epstein, Herman. The electrographic recording technique. *In* Joint Computer Conference. Proceedings. 1955. New York, Institute of Radio Engineers. p. 116–118.
10. Fasana, Paul. An approach to the mechanization of a syndetically integrated subject heading authority file, special report no. 4, Contract AF 19(604)–8438. [Lexington, Mass.] Itek Corp., Information Sciences Laboratory, 1963.
    Another report of this project was published under the title "Automating Cataloging Functions in Conventional Libraries" in *Library Resources and Technical Services*, v. 7, fall 1963: 350–365.
11. General Dynamics/Electronics. *Information Technology Division.* S–C 4020 system description; descriptive material. [n.d.] 13 p.
12. Higonnet, Rene, *and* Louis Moyroud, assignors to Graphic Arts Research Foundation, Inc., Cambridge. Mass. Photo composing machines. U.S. Patent 2,790,362. Patented Apr. 30, 1957; filed Aug. 23, 1947.
13. Higonnet, Rene, *and* Louis Moyroud, assignors to Graphic Arts Research Foundation, Inc., Cambridge, Mass. Type composing apparatus. U.S. Patent 2,951,428. Patented Sept. 6, 1960; filed Aug. 22, 1957.
14. Hoffman, John H. Computerized typesetting systems. New York, ANPA Research Institute, 1963. 5 p. (American Newspaper Publishers Association. Research Institute. R.I. bulletin no. 787)
15. Hosken, J. C. Survey of mechanical printers. *In* Joint Computer Conference. Proceedings. v. 2; 1952. New York, American Institute of Electrical Engineers. p. 106–112.
16. Information Dynamics Corp. A compilation of data on computer-output printers proposed for MEDLARS; technical report to the National Library of Medicine. Wakefield, Mass., 1961. 14 p.
17. Information Dynamics Corp. Costs and material handling problems in miniaturizing 100,000 volumes of bound periodicals. Engineering report # CLR 171 for the Council on Library Resources. Wakefield, Mass., 1961. 30 p. 13 fold. charts.
18. International Business Machines Corp. Graphic composing techniques; report on contract AF 30 (602)–2527. Yorktown Heights, N. Y., Thomas J. Watson Research Center, 1962. 67 p.
    Report no. RADC–TR–61–310.
    ASTIA document number AD 273 414.
19. International Business Machines Corp. IBM 1403 printer: original equipment manufacturer's information; reference manual. New York, 1961. 16 p.
20. Ledley, Robert S., *and* James B. Wilson. Investigation of the use of digital electronic computers in the publication of the *Index Medicus*. Final report on National Library of Medicine Contract SAph 712151. [Washington] School of Engineering, George Washington University [n.d.] 116 p.
    Unpublished report.
21. Markus, John. State of the art of published indexes. American documenation, v. 13, Jan. 1962: 15–30.
22. Mergenthaler Linotype Co. Mergenthaler news for release. Brooklyn, 1962. 2 p.
23. Mooers, Calvin N. The tape typewriter plan: a method for cooperation in documentation. Aslib proceedings, v. 12, May 1960: 277–291.
24. Moore, J. K., *and* Marvin Kronenberg. Generating high-quality characters and symbols. Electronics, v. 33, June 10, 1960: 55–59.
25. McNaney, Joseph T. Electron gun operates high speed printer. Electronics, v. 31, Sept. 26, 1958: 74–77.
26. New computer program automatically prepares data and produces offset printing plates for publication. The NBS standard; official employee bulletin, v. 7, Sept. 1962: 1.
    This news article describes a programming technique for dealing with tabular data. For a full

report of this technique, see *Phototypesetting of Computer Output; an Example Using Tabular Data,* by William R. Bozeman (Washington, National Bureau of Standards, 1963. 6 p. (U.S. National Bureau of Standards. Technical note 170))

27. Photon Corp. The ZIP, a high speed print-out device; descriptive brochure. Wilmington, Mass. [n.d.] 2 p.
28. Photon Corp. The application of Photon phototypesetting equipment to electronic data processing; descriptive brochure. Wilmington, Mass., 1963. 11 p.
29. Radio Corporation of America. Newspaper type justification seminar [West Palm Beach, Florida, January, 1963]—abstracts. [n.p., n.d.]
    Unpublished paper.
30. Segel, Ronald, Tirey Abbott, *Jr., and* Jerrold Seehof. High speed printer survey; part 1. Dayton, Ohio, National Cash Register Co., 1958. 43 p.
31. Smith, H. O. The future import of photocomposition. *In* Association of Printing Technologists. Printing technology; proceedings of the conference, v. 5, Nov. 1961: 7–22.
32. Sparks, David E. A machine interpretable format for library cataloging. Wakefield, Mass., Information Dynamics Corp. [1962] 18 p.
    Unpublished report.
33. Symposium on high-speed printing of computer and tape data. *In* Technical Association of the Graphic Arts. Proceedings. v. 13, 1961. Washington. p. 177–240.
34. Taine, Seymour I. The future of the published index. *In* Institute on Information Storage and Retrieval, *3d, Washington, D.C., 1961.* Machine indexing: progress and problems. Papers presented at the . . . Institute. [Washington] Center for Technology and Administration, School of Government and Public Administration, American University [1961?] p. 144–169.
35. U.S. *National Library of Medicine.* Index mechanization project, July 1, 1958–June 30, 1960. Washington, 1961. 96 p.
    —— —— *In* Medical Library Association. Bulletin, v. 49, no. 1, pt. 2, Jan. 1961: 1–96.
36. U.S. *President's Science Advisory Committee.* Science, government, and information: the responsibilities of the technical community and the Government in the transfer of information; a report. Washington, U.S. Govt. Print. Off., 1963. 52 p.
37. Voigt, Melvin J. Report on serials computer project, University Library and UCSD Computer Center. La Jolla, Calif., University of California, San Diego, 1962. 32 p.
    Processed.
38. Waite, David P. Microfilm card is information medium for space agency. Systems management, v. 3, Nov./Dec. 1962: 27–29.
39. Washington Engineering Services Co. Survey of research on typographical effectiveness, summary report to U.S. Navy Bureau of Ships on Contract no. 6s 76029. Bethesda, Md., 1959. 68 p.
40. West, Ralph E. High speed readout for data processing. Electronics, v. 32, May 29, 1959: 83–85.
41. Yasaki, E. The computer and newsprint. Datamation, v. 9, Mar. 1963: 27–31.

CONFERENCE SESSION V

Output Printing: Introductory Remarks

FRANK B. ROGERS
National Library of Medicine

Sequential Card Systems

We have a very excellent paper from Sparks, Berul, and Waite. There are several types of equipment discussed in the paper. One of these types is designated the sequential card system; there are three commercially available machines which do about the same job. The first one on the market was a Photolist camera. This utilizes a punched card across the top of which a single line of information is typewritten. These cards have filing indicia punched in them; they can be arranged in order by ordinary EAM equipment and passed rapidly through this sequential camera to photograph line by line and set the page for offset reproduction. One of the advantages, perhaps, of the Photolist system is that it *is* a system.

A second type of sequential card machine is the Listomatic. The significant thing about this machine is that one, two, or three lines can be typed across the top of the punched card and the camera will open its aperture, one, two, or three times appropriately for each card, varying this throughout the sequence. This is the only camera of the three that has this capability.

The third type of sequential card machine is the Composolist, which will accept almost any size of card on which information has been typed. The area of the card to be photographed may be selected for any given run. One might have information across the top of the card which could be photographed for one purpose and information in the bottom right corner of the card which might be photographed for another purpose. It is versatile in this way. One must, however, select the area to be photographed at the beginning of the run and this cannot be changed during the course of the run.

These sequential card systems are of interest for library applications. They are not output devices from computers. It is conceivable that they could be operated in some manner by output from a computer, but it is inconceivable to me why anybody would think it worthwhile to do so.

Mechanical Printers

The ordinary type of output device from the computer is the mechanical printer. There are various types of mechanical printers: stick-type printers, wire-type printers, drum printers, and chain printers. Most widely used and of most interest to us are the last two. In the drum printer we have, in effect, a series of disks stacked together in the form of a drum; each disk contains the alphabet and other characters available in the system. Usually there are 160 positions across the drum and one can select 120 of these on which to print out information. The machines typically operate at a speed of 900 lines a minute. The chain printer was introduced by IBM several years ago and embodied in its 1403 Printer. This printer consists of a horizontal, continuously moving chain that has 240 character positions; in the typical application, this chain is divided into 5 sets of 48 characters each. This continuously moving chain is always bringing an appropriate character into the printing position.

The drum printers have 52 or 56 characters available; the typical chain printer has 48 characters. Both are uppercase one-font machines, and the librarian, of course, would be interested in a multiple-font upper and lowercase machine if it were reasonably possible to have this. In our MEDLARS study we investigated the possibility of modifying the drum printers to get upper and lowercase. This could be done, but not very easily since the size of the drum begins to increase enormously and this causes a lot of difficulty. Most interesting is the modification which has been made to the horizontal-chain printe., the 1403 IBM Printer: instead of repeating a 48-character set 5 times around the 240 positions, a 120-character set is repeated twice. In the 120 characters you can get an upper and lowercase font, perhaps one case of another font, and some extra characters according to your needs. Of course, if you are going to use two 120-character sets rather than five 48-character sets, the speed of operation goes down somewhat. While the 48-character setup operates at about 600 lines per minute, I think the 120-character set operates at about 270 lines per minute.

This 120-character set on the 1403 Printer is being utilized at the present time by the Chemical Abstracts Service. In the paper, in figure 16, there is an example of this printing from the modified 1403 which Chemical Abstracts uses for its new publication, *Chemical Biological Activities.*

The paper clearly points out one of the interesting problems in using the mechanical printer. It is the very nature of the machine that you do not have proportional spacing and, of course, you do not have variable leading of the lines. Typically the mechanical line printer prints 10 characters to the inch horizontally and 6 lines to the inch vertically. Now this requires a fairly large amount of space to print a given amount of characters, and it has been shown that in many applications the number of pages to be printed would be doubled by use of the mechanical line printer as opposed to a machine which permits proportional spacing and variable leading. There is possibly one alleviation of this problem which Chemical Abstracts, I believe, intends to use. They thought of the use of the anamorphic lens which compresses the page in a horizontal direction more than it compresses it vertically.

Tape-Operated Printers

Computers can also be made to output perforated paper tape; this means that anything that you can operate with perforated paper tape can be controlled by output from a computer. As the paper points out, one can operate hot-lead composing machines from this perforated-paper-tape output and in some newspaper applications this is being done. One can also operate, with perforated paper tape, some of the mechanical optical photographic composing devices which are typified by Linofilm and by the Photon machine. Linofilm runs from a 15-channel paper tape. (There is a converter available for translating magnetic tape output into this 15-channel paper tape to operate Linofilm.) The Linofilm and the Photon run at the same speed as the perforated-paper-tape typewriter, that is, about 10 characters a second. They have a very wide range of fonts, proportional spacing, variable leading, and right-hand justification. These machines give an excellent typographic product and are in widespread use.

The last class of machines discussed are those which might be operated directly by a magnetic tape output from the computer. At the present time there are two classes of these machines. One is the cathode ray machine in which characters are formed by an electron beam on the face of a cathode ray tube. The rapidity of this device is very great; an enormous number of characters can be formed quickly. One must then get the image off the cathode ray tube with some kind of photographic electrostatic device. These cathode-ray-tube printers are now in use. I believe that some magazines with large numbers of subscribers print their address tabs from cathode-ray-tube devices and then take off the print electrostatically from the tube. There are some severe engineering problems involved with this type of device; the size of the face of the tube is not very large. In the largest of them there is the problem of bending at the edges of the tube, but they are extremely fast.

The second type of device is a variation of the mechanical optical photographic composing machine driven by a magnetic tape; this is typified at present only by the GRACE machine, which is being built by Photon for the National Library of Medicine. The GRACE machine accepts magnetic tape; it has three fonts in upper and lowercase, special characters including some diacritical marks; it has a total set of 226 characters. GRACE operates at a speed of 440 characters a second, composing character by character on film.

The way in which it differs from the ordinary Photon mechanism is, briefly, as follows: In the ordinary Photon device the fonts are set up on a disk which has eight rings. Each ring holds two type fonts, one on half of one side of the disk and one on the other half of the disk. There are 16 fonts on the disk which are accessed by moving from ring to ring as the different fonts are needed. Photon has a single light source which punches through this revolving circular matrix and throws a character on the film. In the GRACE machine there is a single oblong matrix. As far as the machine is concerned it only has, you might say, one font of 226 characters; no font shifts are involved in the operation of the machine. For each character in the font there is a light source which fires through the matrix and the lens travels from left to right across the line as it composes a line and then from right to left as it composes the next line, and so forth.

Now I would like to throw the session open for discussion.

General Discussion

PATRICK: The authors gave us the cost of the mechanical chain printer on the graph in figure 28, but they neglected to include its performance on figure 27 [see pages 165 and 166]. This is extremely important because the chain printer is commercially available to the library society without any further development. I have the curve here if anyone is interested in it. The lower bound of the curve against the extreme right-hand margin is $.0024, a quarter of a penny, per line in the 100,000-line volume, and it gets down to a tenth of a cent per line in the 240,000-line volume. This cost includes the impression and the printing. I would like to reopen the debate of yesterday concerning the number of fonts needed because if you cut the fonts down you can achieve a significant saving in costs.

F. B. ROGERS: It would be interesting for us to know some of the assumptions on which the authors based figures 27 and 28.

BERUL: The primary reason that the chain printer was not included was because figure 27 shows tape-operated photocomposing machines and illustrates one point only, unit cost vs. volume, rather than a comparison of the merits of various systems. Certainly this cost-vs.-volume relationship is true for the chain printer as well. We were trying to emphasize here that if machines such as ZIP or GRACE or VIDIAC, which embody the character generator concept of the cathode ray tube, are operated at low utilizational levels they are going to cost as much or more than the lower speed, 10-character-per-second machine. If they are operated at the high end of the curve, they become efficient, as far as composing machines are concerned, in cost per line composed. This figure shows the effect of production utilization on unit cost and illustrates the fact that these big machines are not for the little user.

In figure 28 we show a curve which goes up at a higher rate for the chain printer than for the more sophisticated graphic arts quality composers. This is due to the increased number of pages which result from using the chain printer with a low type density (10 characters per inch horizontally, 6 lines per inch vertically) even with reduction to achieve some compression. This is also illustrated in figure 17 where we show the same copy composed by two techniques. We do not think that for a high volume—that is a large edition—publication the mechanical printer is a good medium for composing graphic arts quality. If you have a large printing run the effect of that printing run on cost is going to be significant because you're going to print twice as many pages. Now there are some controversies as to whether it is going to be 60 percent more or 140 percent more. Some competent studies, e.g. the MEDLARS study, used the figure of 100 percent; and I have checked it out and it's correct in their particular application. This comparison really depends on what your exact system requirements are going to be.

PATRICK: One point on figure 26. Some people may not realize that figure 26a has 36 waste character spaces embedded in white space in the text. On the normal printer it happens that if I take these out (this might require only a few months' experience beyond the first formatting) I can print them two up, double the output rate, and halve the page size. To achieve a fairly good quality, I can lay them photographically with a format already on the sheet and run them on lined paper, achieving in this particular case almost the same font as 26b but in columns and at maybe a tenth of the cost.

BERUL: If you loot at footnote 29 and the sentence to which it refers on page 163, we reported that it would take 7½ pages to compose the same amount of material by the graphic arts quality printer compared to 20½ pages for computer printout. I would agree that if you are going to print 1 to 5, even 100, 500, or 1,000 copies, as far as cost is concerned the computer output printer may be cheaper. It is in large runs of 5,000 or 6,000 copies that the costs begin to hurt. Now, this ignores the quality considerations. There are other reasons for having multiple fonts. We tried to show what it looks like and point out the differences, but this is a purely subjective matter.

CLAPP: I consider the topic of this meeting possibly the most important of the conference from the point of view of opening the doors of library work to computer applications. Until there is a decent font available, it doesn't pay anyone to put a great deal of bibliographical material into machine-readable form. Until there are large quantities of bibliographical material available in machine-readable form, it doesn't pay many libraries to engage in machine processing. Until MEDLARS, there has been no real possibility of widespread library use because there has been no real development for getting a decent font of type.

Patrick raises the question as to what this font of type should contain. How large must it be? How small may it be? No one has really studied this, Mr. Patrick, except the National Library of Medicine from its particular point of view. I'd like Brad Rogers to comment on this and respond from the point of view of the National Library of Medicine, which does have a very heavy bibliographical load which it has to transmute into graphic form through the machine-processing systems. I foresee the probability that the Library of Congress will soon be called upon to produce machine-readable bibliographical information. However, before we can expect the Library of Congress to do this, we have to come to some agreement as to what this font of type is and this will take our best brains. Meanwhile, we could very well learn from what the National Library of Medicine has done up to this point.

F. B. ROGERS: The main product that we wish to print with GRACE is the *Index Medicus*, a subject index to the periodical literature of medicine. We need enough fonts to distinguish the subject headings from the citations and to distinguish, within the citations, the beginning of the journal title abbreviation from the author and title. We have a 6-point font, which is a quite small size, probably the limit of what you would want, for the citation itself. We are trying to pack 10,000 characters on a page. Our present load is 5 to 6 million characters a month and will soon be 10 million characters. It costs a lot of money to print that many characters; therefore, we have adopted the smallest type size that we think is at all reasonable.

CLAPP: What is the size of the font?

F. B. ROGERS: The total character set is 226. There are 3 fonts upper and lowercase included, within that total of 226 characters. There are Greek letters, but no Cyrillic characters, in the set.

CLAPP: Are there mathematical symbols?

F. B. ROGERS: No, there are no mathematical symbols.

PATRICK: What is the speed of GRACE and its initial cost?

F. B. ROGERS: The speed is 440 characters a second. The cost of the first device is somewhat in the neighborhod of $300,000.

CLAPP: What would be your guess as to the adequacy of these fonts for general bibliographic use?

F. B. ROGERS: I think it would be entirely adequate. One of these is a 10-point font, one is a 6-point font, one is boldface. I don't see why this would not be entirely adequate for bibliographic purposes.

SPARKS: I might point out here that general bibliographical use is a rather undefined thing. Someone yesterday remarked that the small library gets along with a typewriter in spite of the fact that the Library of Congress cards are printed in about four fonts! This is true. So we can see that there is a bibliographic use of a minor importance, or a less demanding bibliographical use, which would perhaps require only a single font. A tape-punching typewriter would satisfy this need. But when you mass bibliographical citations on a page and reduce their size, you place demands on the human eye which must be taken care of by providing a variety of type sizes and a flexibility in placing the images of the characters in the proper place. This includes reducing the space between lines and reducing the space between characters to form a psychologically acceptable document. I would say that an acceptable font for bibliographical use would have to define various sets of objectives. Some are more important and more demanding than others.

WARHEIT: One thing I do want to mention about the chain printer for the benefit of librarians here. There is another device on that chain printer that's called a "90-degree rotation" and, if any of you have been getting catalog cards printed off the 1403 Printer, you should ask the salesman about that 90-degree rotated chain. On the standard printout you will get about 17 lines on a 3 by 5 card; with the rotated chain you will get 25 lines and faster output.

DUBESTER: I think Patrick's comments and questions deserve attention. We know that libraries do not rely completely on LC cards. Now what about the cards which these libraries produce themselves? What I have seen includes order slips, photocopies thrown into the catalog, typed cards, and printed cards with higher or lower quality. In other words, what the libraries will accept is a very proper question. I think that what libraries will accept will be different for a card catalog than it will be for a book catalog, since these have a different type of use. For example, 6-point type is too small for a catalog card which has to be looked at for any length of time, whereas in a book catalog where a 6-point type signals a certain kind of information (or, more properly, signals information that you can skip until you find the particular item which requires you to look at it) it can have virtue. In other words there is an objective, a performance aspect which has not yet been properly analyzed in terms of these type variances. The human engineering function in catalog use has perhaps been least studied of all in our library operations.

PATRICK: It seems from the limited library work that I have done as an engineer and researcher that we are all very willing to demand high quality of LC cards, and we will spend every last dollar that LC will put into those cards and insist on perfect quality, because it doesn't come out of our pockets. In the libraries I use, we seem to get along quite well with material that's not as high quality because it is produced locally. I have some cards in my briefcase that show that some people are even getting along with 407's, straight output, one type font, no proportional spacing; the product is not beautiful, but it seems to work. I really think you ought to discuss this because it is very important in determining how rapidly you can move. We could give you catalog cards overnight, printed in quantity, if you could get along with a little lower quality—not 6 months, or a month, but 24-hour service in the mail. The catalog information would be printed on one side, your address printed on the other side, the sides folded together like a utility bill, and then the cards become a post card.

ORNE: The problem we are talking about has relationship only to the largest libraries in the country. The quality of the LC card is determined first by the needs of LC. It always has been. The outfall for other major libraries is fine, but actually the commercial possibilities of what we are talking about today will reach only a very few of us and has little importance except for the point of view, as Verner Clapp pointed out, that it may open up other possibilities.

DIX: The thing which strikes me as most interesting here is this concept of a store of bibliographic information and a printout to order. Is the potential cost of automated printing of a copy-to-order single set of cards at all within the price range of edition printing of that card when the cost of the storage, maintenance of stock, and all the other expense is considered? In other words, is there any possibility in the future of a mechanized copy-to-order process by which the card will be printed only when one orders it, or is this out of the realm of discussion?

F. B. ROGERS: I think it is beyond reason, but let's hear what Patrick has to say.

PATRICK: I would like to start with commercially available equipment which we could use now if we had the file converted. I would like to describe for you the file converted and stored on disks; at present it would cost about $5½ million just to retain the file. This file would be equivalent to the National Union Catalog; it would be up to date to the 24th hour: it would be updated every night. There would be several searchable files on this. You could write in, as if you were ordering parts out of a warehouse, and give the numbers or the identification of each card you wanted. In addition to this, we could have on file your classification scheme. We would then print your cards on these printers we have been talking about, if you can get along with just 64 characters, because these are all we have today. The picking price, fetching them out of the file, would be quite normal, something like a tenth of a cent per card. The cost of printing would be a tenth of a cent per line on the card; the cards would come to you in order ready for filing. For the average card, which has about 8 or 9 lines on it, this would cost a penny printed on the 1403 and in the order for interfiling. Today this can be done competitively if you chop the quality.

WAITE: It can be done competitively under certain circumstances which are complex and which take a great deal of time to study carefully Dix raised a question about the feasibility of output printing on demand. I think there is a possibility for such a service with output printers of the kind described in our paper, but it would have to be on very high-production equipment to get the unit cost down very low in order to compete. What you are asking for is a copy of some graphic image, actually, and there is no need to go through all the coding and uncoding paraphernalia in order to reconstruct it. As far as physical handling is concerned, we would have to get into quite an elaborate study to come up with an exact answer. I think, however, that the promise is perhaps more on printing graphically than digitally for on-demand printing.

I would like to quarrel with Orne's statement that the subject of high quality in the output printing products of libraries is of concern only to large libraries and specifically to the Library of Congress. The missing element in the discussion so far is that the librarians have not stated the output products and services which they wish to provide for the user community. When they do, the technical people and the systems engineers can start working and can begin to play the game of alternatives. One of the things that will happen is that output printing for special patron communities and for the small libraries will be a distinct possibility.

WARHEIT: We have been focusing on the catalog card. I agree that demand publication from a graphic image will be and is now cheaper. However, on the other hand, the digital index, which would be your catalog, has other bibliographic uses. Its purpose is not just the making of catalog cards, but also the searching of that same index for printing out special bibliographies. We should consider the products the librarian wants in order to determine, as you say, what alternatives should be selected. If it is just demand printing, I agree that graphic output from a graphic store, where there is no conversion problem, will be the most economical. But the librarians really want a number of end products. Once they get one catalog card or one set of cards they are through with that, but they are going to use the store over and over again for other purposes such as announcement bulletins, reading lists, searches, and the output of the searches in response to reference requests.

NAESETH: I plead very strongly for, not necessarily a wide variety of fonts, but legibility, because whether it's catalog cards, lists, or anything else, we want these easily read. I feel strongly that the present IBM font is just not satisfactory for us. I am surprised that no one has mentioned the unusual fonts: Slavic, Greek, Indic, and so forth. I believe we could get along without most of those, except perhaps the Greek. We can get along fairly well with transliterated Russian, and the Library of Congress gets us cooperative copy for Indic languages in Romanized type. We get along fairly well with that, even though it may be a little more offensive to the users in the vernacular. So with Romanized type and perhaps with a little bit of Greek we could get along, but I plug again most strongly for easily legible print, which first of all means upper and lowercase and a better font now than we get from IBM.

ESTERQUEST: There is one further consideration with respect the integrity of the language, even within the Roman alphabet. Most of the librarians here represent educational institutions where, for example, Polish and Danish are taught in the classroom, and the student is told that this word is "k-r-ø-l" with the slash going through the "o." It seems that before librarians retreat too much on this matter of quality, we ought to recognize that we have an obligation to maintain the integrity of these languages. This argues for not giving up even the more exotic diacritical marks.

BERUL: Dr. Rogers indicated that he thought that the GRACE machine would be adequate for most bibliographical requirements. I agree, but one point was left a little unclear about the character set. I do not agree that 226 characters would be adequate as a total set for all requirements; I have studied this problem and I found that in many places, including libraries, a 1,000-character set may be required. If you ever walked into the Government Printing Office branch at the Library of Congress where they are composing catalog cards you would be quite impressed by the dexterity and cleverness of the Lintoype operators who are composing in multilanguages. The Linotype font has fewer than 226 characters: it is interchangeable. This is true also of the font in GRACE. The entire font can be removed, put aside, and, in a few seconds replaced by a new one. Even on a Photon machine which has 1,440 basic characters, the glass can be unscrewed in about 10 seconds, put aside, and a new font inserted. I believe this can be done also with ZIP and with the chain printers. It's a question of identifying the particular parameters of your system and separating these problems, as is done now with LC catalog card composition. Composition in Russian is given to the person who is typing on the Linotype machine for Russian. If he has to do some Arabic on the same machine, he will put in the magazine for Arabic or go to another keyboard. This problem could be attacked by segregating the work into groups; you could compose Arabic, then Cyrillic, and so forth.

If I may answer one other question. I agree with Patrick—we also used three-tenths of a cent, per line as the cost for mechanical output printing, as compared to around 4 cents for photocomposition. The 0.3 cent figure for the 1403 printout allowed a generous amount of efficiency at a 300-minute day instead of a 400-minute day (coffee breaks and the like). (Even though they say that the machines don't take coffee breaks, they do every once in a while.)

In answer to the question about simple devices for small libraries, we stressed the concept of a wide range of output printing problems from the simplest problem to the most complex, from the smallest volume to the largest volume application. We follow, at least I do, the same philosophy that Rutherford Rogers so eloquently expressed last night when he stated that bibliographic control is best done centrally so that it will not be duplicated all over the country. We assumed that with computers this problem would be attacked using centralized production techniques. Therefore we assumed that by having this great mass of bibliographical material in a machine manipulatable form, you could achieve an output product that could service the entire nation.

An example mentioned in several papers is the National Union Catalog of 14 million catalog cards. Actually this catalog already avoids duplication because the code number for a particular library is posted on a master card to indicate that it has a particular title. Thus, one master card records all the libraries holding that particular title. So with respect to publishing the National Union Catalog one has to think of the most sophisticated output printing devices. For the small library, which wants to produce a set of catalog cards for what it is cataloging now but which also wants to contribute its share in this massive task of bibliographic processing, we have suggested that their input to these massive stores can be solved by the simple device of the tape typewriter, or even the keypunch machine, where there can be a byproduct of a manipulatable machine-interpretable record. This record can be used in combination with computers or with special purpose devices, such as the Itek Crossfiler and with the Selectadata, where the tape is used to automatically produce a full set of catalog cards with all the headings overprinted.

I agree with Patrick that it is also feasible to query a central store and ask for a set of catalog cards. I am sure this can be done for three-tenths of a cent a line or one-tenth of a cent a line if you get really high-production loading efficiency. It may cost a penny to get that card out of the ma-

735-898 O—64——14

chine, but it will probably still cost you a dollar to order it.

F. B. ROGERS: I have two comments to make. It is true that you can take this 226-character set out of the GRACE machine and put another 226-character set in. This will take time; it also means that Indic scripts and Roman scripts cannot be used at the same time and that's what we would like to do. I just want to point out that you have to pay for everything and you have to decide what is most important to pay for. To get a speed of 440 characters a second on this machine we not only give up a large number of fonts that are available on the slower speed machine, but we also give up other things. It is not nearly as easy, for example, to change column widths on this machine as it is on the slower speed machines. Also on the high-speed machine you have only a manual setting for vertical leading; there are not the tremendous possibilities of changes within a run that are possible with the smaller machine. So it's just a question again of how much you want. What is the priority? What can you pay for it? What is most important?

Just a minor point, the GRACE machine was referred to in the paper and by the Photon people as the ZIP machine. This is sometimes confusing, because the name ZIP was originally used 2 or 3 years ago, also by Photon, for another machine which they had in contemplation—a paper-tape-driven device on which the speed was about 100 characters a second. So if you just understand that there was one ZIP 3 years ago and there is another one called ZIP now, no confusion need result.

ELLSWORTH: The cost of the cards that would come out of the machine is not important in itself. It is important only when it is related to what happens in the libraries that use the cards. Now this is a problem that seems very simple to anyone who has worked on it; there are a dozen librarians in this room who have worked on the possibility of centralized cataloging for a long time. We know perfectly well that if all research libraries would do certain things and operate in certain ways in relation to the Library of Congress, it wouldn't matter if the cards that came to them cost $1, $2, $3, or even $5 because it now costs them a lot more than that to catalog in each library. We know that the situation could be changed without any machines and that radical economies would be made, but we have not been able to persuade the profession to do this because basically I think they don't understand it. I think that it would be important for those of you who are not librarians to remember that the cost of the card in itself is not really important, except in relation to what we would do to the rest of the system in our own shops.

ANGELL: I believe the group might be interested in a brief report on an experiment on this matter of output printing that we have conducted in the Subject Cataloging Division in the Library of Congress for the past few months with the help of Ed Forbes at the Government Printing Office. Rutherford Rogers alluded this morning in his review to the problems that we have in updating and publishing our subject heading list. This is a compilation of the subject headings which are used on our printed cards and widely adopted by a number of libraries. The sixth edition of this list is a publication of some 1,137 pages; it appeared 6 years ago. Our period for basic cumulations is much too long; we keep it up, as many of you know, by monthly and annual supplements. Now we were told by GPO, in designing the sixth edition, that the production of the cumulated volume is the most difficult technical printing production in that plant. This gives the Subject Cataloging Division a certain distinction, but one of which it is very anxious to divest itself. A few months ago we gave Forbes sample pages from the sixth edition and a set of sheets as though we were supplying the changes that would need to go into the seventh edition, so that he could put it on tape and wave the various wands that are required in graphic production and simulate the first pages of the seventh. We prepared a code for broad classes, just the first line of our classification, in order to test in a primitive way the possibility of extractions of special subject heading lists, which is one of the things for which we are importuned by special libraries.

The first results appear to us to be extremely promising. There is no loss of the very great typographical sophistication that the list has given us. We are quite hopeful that this will be a feasible operation. Now there is one thing that Forbes doesn't know about, and that is that we're also importuned to publish our subject headings on cards. Partly this demand comes from the desire for updating, but it also has an independent impulse. If this works, we will try it out on the classification schedules: those are instruments of relative notation reputed by some to have a certain utility in the management of library collections. I look forward to the time when we can say that we have done this.

SNYDER: I would like to go back to character counts. Peter Brown at the British Museum promised to send me the results of a 7-month character-count study of the British Museum cataloging output. I have encouraged him to publish this although he didn't feel at the time that it would be useful to enough people.

At MIT we also were making character counts in the processing of cards in our catalog department. The long-range viewpoint of this count is focused on catalog card reproduction by typewriter. We can use LC cards for about 33 percent of our material. We have a considerable quantity of cards for which we can use the typewriter with satisfactory results. If we can determine the number of characters necessary, and we hope to be able to do this, we will be able to know at what point we cut off, 88 characters, 94 characters, etc. I feel, at this point, it's not the 94 that we suspected. The IBM Selectric does give us a little bit of type font versatility. We could batch our material by language, and if we can get within an 88-character keyboard, we can do something like 92 percent of our material.

VOIGT: I would like to raise a question which I think has only been alluded to here today, although it is mentioned in the paper. Could you give us any indication at this time of what are, or what will be, the best methods for printing from computers when we want more than 1 copy but not hundreds or thousands of copies, when we want 2 to 5 or perhaps 10 to 20 copies? This is a real problem to those of us who have started working with computers in our operations. We want to mechanize our serials lists and author-title catalogs which we would like to reproduce for the use of our readers at various locations in our libraries. We can see the possibility of using the computer but thus far we have not had notable success, in our case, in duplicating these in small quantities.

PATRICK: We quite frequently run vellum for short-run, very cheap copies, and go through a blue-line machine which makes either blue lines or black lines. Also there is a process, rated at 600 lines a minute but about 450 lines a minute net, where we print directly on Multilith masters. You can get, from fanfold paper mats, 600 to 1,000 very good copies, even using both sides.

FORBES: I would like to talk a little about long-range planning at GPO. We have a conviction that very clearly there is a far greater need for typographic composition at very high speeds than even we thought was possible. The point is perhaps best illustrated by the magnitude of the output printing problem in the defense establishment. In one operation alone there was a requirement for about 2 billion printed pages a year from information already standing on magnetic tapes.

I would like to comment further about these high-speed typesetters. The technology is such now that character generation of a variety of type fonts electronically has been demonstrated and is quite practical. The significant thing is that complex graphic arts quality symbol generation is going to be reasonable in cost, and it will obviously be possible to parallel many graphic symbols on machine control so that they can be displayed on the face of a very high-quality cathode ray tube or television tube. This has some very far-reaching implications, and I believe it removes a lot of the problems you are concerned about with respect to the nuances of typography needed on catalog cards to communicate the information with clarity.

Just to see if it would be humanly possible to compose typographically Library of Congress catalog cards on demand by machine, I took 50 million cards (this is slightly more than present card sales) and divided them by 250 working days in the year. This requires roughly 200,000 cards to be set each day. The raw material for a card would cost about 2 cents. Now if you could utilize a high-speed typesetter to do this job (I figure 10 lines to a card and less than 100 characters per line) you need a capacity of something less than 1,000 characters a second. We know machines like this are very practical technically. This would mean that the requirement could actually be met by 70 machine-hours a day on high-speed phototypesetters.

You can visualize then that if GPO, for example, had several such machines directly connected to the Library of Congress and if there were any interest in offering a service such as this, three high-speed typesetters could do the job. Now this is just an academic exercise, but I think calculations like this are helpful to give you some perspective on what the technology is going to do within a few years.

There is something fundamental to be learned from the 1403 Printer and from similar machines as we see them today. We all recognize that 10 years ago off-line, high-speed printing was a very important part of a computer system operation. At that time, a number of machines which were essentially magnetic-tape-to-control-unit-to-electromechanical-printer systems were built at a cost of roughly $200,000. The real limitation of these machines was the fact that they were never utilized enough to absorb the high cost of the control electronics. The 1403 system has been so successful because for about $5,000 it combines a number of functions and uses a common set of sophisticated electronics, so that you can use the machine as a card reader and control, as a magnetic-tape unit and computing unit, and, of course, as an output printer which works simultaneously in these different modes. This has had a tremendous impact; about 5,000 high-speed printers have been installed.

The thing that we see then, from the long-range point of view, is the fact that the electromechanical portion of the high-speed printer system is about a $35,000 device when it's mass produced and tacked on to a general purpose computer. A phototypesetting mechanism is not too much different from that. It's my personal opinion that we will see high-speed typographic printers functioning in the next generation of small-scale computers. If the cost of the machine can be brought down so that high-quality typesetting can be possible at essentially the same cost as present tabulator quality, there will be tremendous utilization of such facilities.

It has always seemed to me that every job on a computer which results in a report being generated is essentially a job that was worth doing in the first place and it doesn't really bother me much whether 1 person or 10 or a 1,000 are going to read that report. If it was important enough to do it's important to communicate the results of the computer's manipulations with precision and clarity. Now if you could do it for practically the same cost and have half the bulk, you would prefer typographic composition.

WAITE: I'm looking forward to the LC automation report as being a guiding light on what the objectives, the products, and the services of the library community are. That's the starting point, as far as I can see.

SPARKS: We have talked about output printing and about library cards, but I think we still need to find out from the librarians what they want to print. Do you want to print your entire catalog in book form? Do you want to print subsets of the catalog in book form? Do you want to print your classification schedules by machine? Do you want to print your subject heading lists by machine? When you have defined what you want to print, then you must look at each one of these publications and describe its demands for typography. You must describe each publication not only in terms of the typography now used but also in terms of what you can sacrifice for machine composition. When you have done this, you will have a set of requirements for machine composition.

BERUL: I would like to add that, depending upon your requirements, you may not need a computer system. Forbes just mentioned that the electromechanical portion of the computer output printer would cost about $35,000. The same thing is true about some of the photocomposing machines; if you don't have them hooked up to computers or need the real sophisticated system, they may be a little cheaper too.

So the point in summary is that computers are not *necessary* for printing; they may be useful; they may be helpful. Smaller libraries with modest output printing requirements or without manipulation requirements may use several devices for input composing, whether they be typewriter or Photon depending on taste or need for quality, and the bibliographical data can be captured at the source in machine-interpretable form. This may be used at some future time for a centralized biliographic record.

R. D. ROGERS: I want to thank the discussion leaders for a first-rate job and all of you for your contributions from the floor.

SECTION VI

Library Communications Networks

Library Communications

J. W. EMLING, J. R. HARRIS
Bell Telephone Laboratories, Inc.

H. J. McMAINS
American Telephone and Telegraph Co.

Introduction to Electrical Communications

Before examining the role of electrical communications in library mechanization, we propose to review communications broadly to provide the uninitiated reader with the background and the special vocabulary which he will find useful in the more specialized portion of the paper. The more knowing reader will hopefully bear with us or skip to the next major section.

We have all become accustomed to three main forms of communication: the spoken word, the written word, and pictures. Electrical communication provides for transmitting all three. In addition, in recent years it has become necessary to provide communication with and between machines as well as between people. Fortunately, the kind of symbolic communication or telegraphy used for the electrical transmission of the written word is applicable to machine communication as well, so we need still consider only three forms of electrical communication:

1. Voice
2. Symbolic, or more specifically, digital signals
3. Pictures

Voice Transmission.—This is by far the most common form of electrical communication and is worth some detailed examination even in a discussion of communication between machines. The pathways or channels for speech provide a ready means for transmitting a useful amount of digital information or, used for voice transmission, they may serve as useful connections between the ultimate user of the machine and a human intermediary at the man-machine interface. Our common experience with voice communication has an important influence on what we expect from other forms of communication.

For perfection voice communication requires the transmission of a band of frequencies from about 40 cycles per second (cps) to about 10,000 cps. But a very satisfactory grade of communication can be achieved with a much restricted band, and commercial telephony employs the frequencies from about 300 to 3,300 cps. Electrical voice communication is commonly accomplished by transmitting an electrical wave that is a replica of the speech wave in air (except for the effects of band limitation), and hence this is referred to as analog transmission.

Speech is a very redundant and inefficient way to transmit intelligence. It can handle speeds of not much more than 200 words a minute and the information rate is something less than 25 bits per second. (Bits per second will be discussed below.) However, speech has some other important characteristics. A listener familiar with a talker can recognize him from his individual speech characteristics. In addition, a listener will promptly note uncertainties in the message (whether due to talker, listener, or transmission system) and can quickly ask for a clarification. Thus there is not only means for identification but a built-in error detection and correction mechanism which we have come to rely on very heavily. Moreover, talkers and listeners are accustomed to adapting themselves to and compensating for variations in speech to an unusual extent.

Digital Transmission.—The 26 letters of the English alphabet have proved a highly satisfactory set of symbols for recording speech. But for the electrical transmission of the written word it was found desirable to use far fewer types of symbols, and a coding scheme was introduced in the early days of telegraphy based on using only two symbols of a very elementary form. Originally these were short (dot) and long (dash) impulses of energy assembled in various combinations to form the characters of the alphabet. Later it was noted that if the individual impulses were sent at regular intervals it was not necessary to have two kinds of impulse. It was only necessary to note whether the impulse at the appropriate instant was present (on) or absent (off). In telegraph parlance they are known as mark (on) and space (off) signals. Each character of the alphabet could then be made up of a series of mark-space pulses in various combinations as in the Baudot code (fig. 41A), long used with printing telegraph machines. These mark-space signals also represent the 2 digits, 1 and 0, of the very simple binary arithmetic used by digital computers, and the information conveyed by a single pulse (on or off) has come to be known as the bit, a contraction for binary digit. This symbolic, or digital, transmission (or some variant) is the most common way today to transmit the written word and to send information (data) between machines.

Just as the printing telegraph uses a group of bits as a code to represent an alphabetical character, machine communication also uses a group of bits as a code, and by analogy these codes also are called characters. Since it is often desirable to use more characters than can be obtained with the 5-digit Baudot code, it is common to use more bits per character, and recently it has been proposed to use a 7-bit code as a standard for all kinds of information exchange, including teletypewriter. This proposed ASCII code (American Standard Code for Information Interchange) is illustrated in figure 41B. The bits in a character may all be sent simultaneously by separate paths (parallel transmission) or they may be sent in succession (serial transmission).

Regardless of the means of transmission, the total number of bits transmitted is a measure of

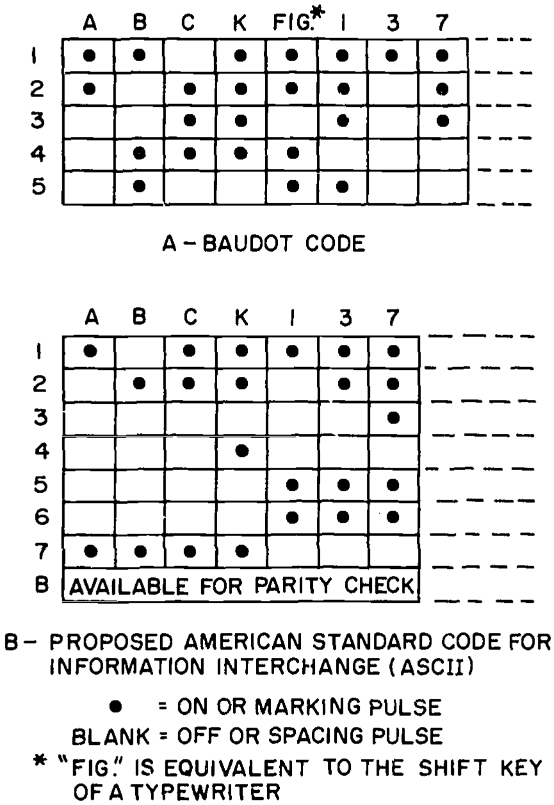


FIGURE 41. *Codes for information interchange.*

the maximum amount of information that can be contained in the message, and the rate in bits per second (bps) is a measure of the maximum speed at which information can be conveyed.[32] To orient the reader, it may be helpful to explain that a 100-word-per-minute printing telegraph machine requires transmission at the rate of about 110 bps in the proposed ASCII. On the other hand, real-time communication between computers may be at the rate of 100,000 bps or more.

It should be noted that digital transmission is communication reduced to its basic elements. It is no longer necessary to transmit a close replica of the original signal as in analog transmission. It is only necessary to have a signal which can be recognized as the presence or absence of a pulse. So long as this can be accomplished, it is possible

[32] Strictly speaking, the actual information conveyed may be considerably less than indicated by the bits transmitted because of inefficiencies in the use of the digits.

to recreate or regenerate the original signal since this consists only of the on-off pulse (or the 1 and 0 of binary arithmetic). Thus it becomes much easier to cope with the signal deterioration which accompanies transmission over long distances. Instead of limiting this degradation through system design, it is only necessary to regenerate the signal before it has deteriorated beyond recognition, and it is then ready to travel farther. However, in this elemental form the signal no longer has those special attributes of speech: indentification and error detection and correction. These benefits, if required, must be obtained by the transmission of additional information.

Picture Transmission.—The electrical transmission of graphical material is ordinarily accomplished by breaking up the picture into a succession of parallel lines. An electrical wave is generated for each line, varying in magnitude with the intensity of light from the line as it is scanned from one end to the other. Ordinarily the electrical wave analogs for the various lines are transmitted in succession and when translated at the receiving end into variable light intensity the picture can be reconstituted (fig. 42).

The transitions of intensity along a scanning line may be very gradual or rather sharp, as when a highlight is adjacent to the blackest part of the picture. Hence, all amplitudes of signal within the range between the lightest and the darkest part of the picture must be transmitted if all tones between black and white are required.

An important class of pictures of particular interest to librarians may have only two amplitudes: white or black. Examples are reproductions of printed or typed pages and simple line drawings. This type of picture transmission is sometimes called facsimile, or Fax, to distinguish it from pictures with tonal gradation: we will use this convention.

It is obvious that the amount of detail (or resolution of the picture) transmitted depends on the number of scanning lines and on the distance along the line in which the transition can be made between black and white. This latter characteristic can be expressed as an equivalent number of lines at right angles to the scanning lines, much as if the picture were divided into small square areas (sometimes called picture elements). For optimum use of the picture elements it is desirable to have a slightly greater number of scanning lines per inch than picture elements per inch along the scanning line. The product of scanning lines times picture elements per line is related to the total number of bits of information conveyed. The relation is about one-to-one for one form of transmission, and this relation is used here in the interest of simplicity. There are, however, forms of transmission which require several times as many bits per element. There are also band compression or encoding schemes which can reduce the number of bits by a factor of four or more.

Figure 43 shows the way in which the number of lines per inch affects the resolution or definition of simple, black-and-white facsimile transmission. For commercial facsimile transmission, a resolution of 96 scanning lines and 67 horizontal elements per inch has been standardized. This gives an image slightly better than the best resolution shown in figure 43. This resolution is adequate for pica type but would not be sufficient for the smaller fonts sometimes used on catalog cards.

If we assume that typing covers about 80 percent of a page, an 8½ x 11 page would require the transmission of about 500,000 bits at 96 lines per inch unless band compression techniques are used. In principle single picture transmission can be sent at any speed desired. The page discussed, if sent at the rate of one per minute, would require elec-

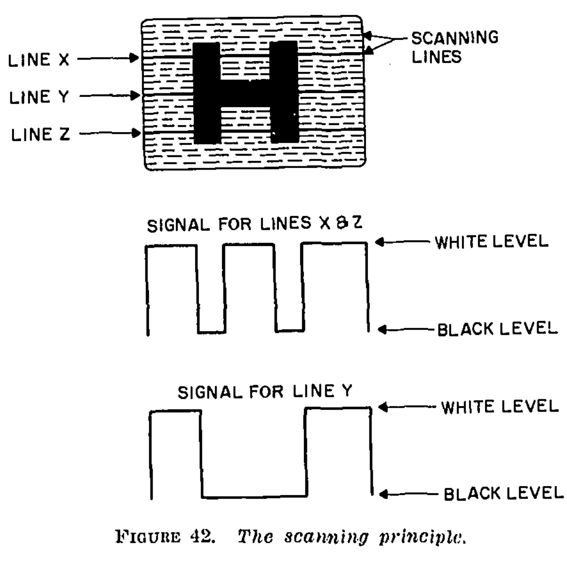
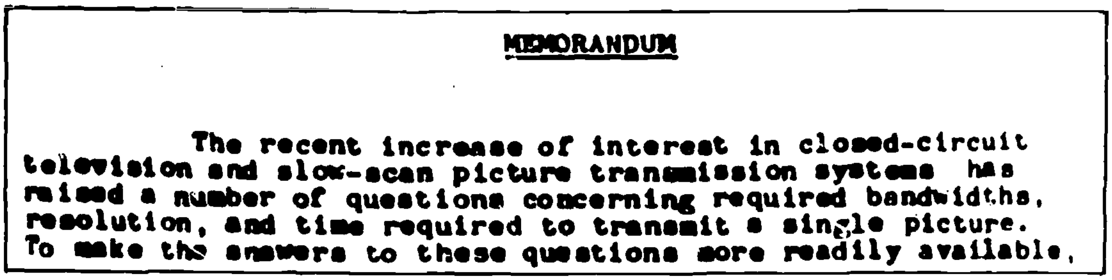


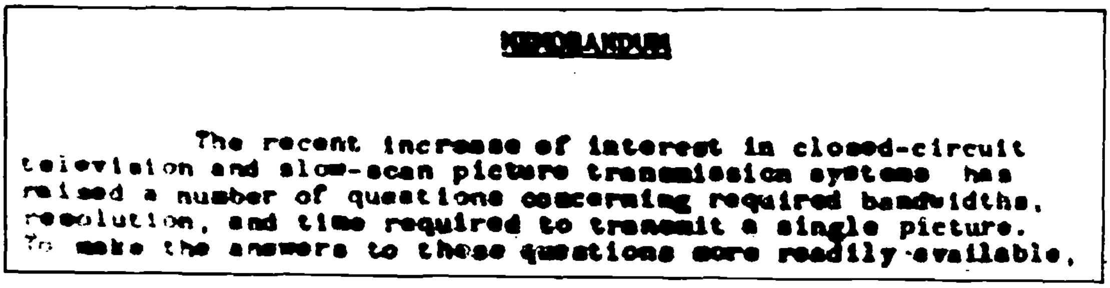
FIGURE 42. *The scanning principle.*



MEMORANDUM

The recent increase of interest in closed-circuit television and slow-scan picture transmission systems has raised a number of questions concerning required bandwidths, resolution, and time required to transmit a single picture. To make the answers to these questions more readily available,

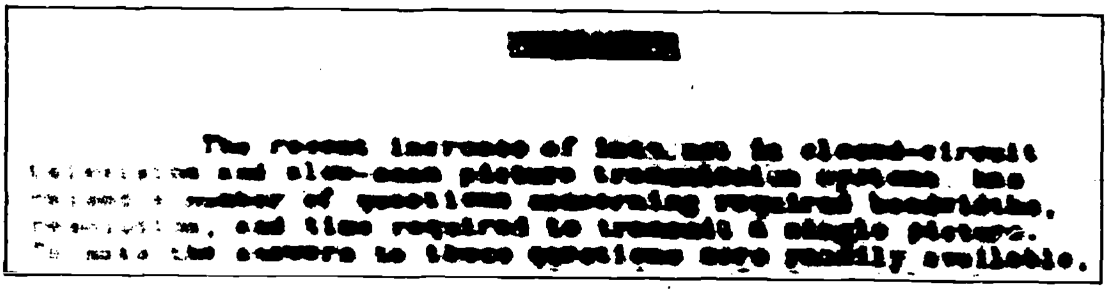
87 LINES PER INCH



MEMORANDUM

The recent increase of interest in closed-circuit television and slow-scan picture transmission systems has raised a number of questions concerning required bandwidths, resolution, and time required to transmit a single picture. To make the answers to these questions more readily available,

54 LINES PER INCH



36 LINES PER INCH

FIGURE 43. *Picture quality vs. number of scanning lines.*

trical transmission at the rate of about 8,000 bps. In order to convey motion, as in television, and at the same time avoid flicker, it is necessary to transmit 25 to 30 pictures (frames) per second. At this frame rate, the bit rate for the high resolution page would be the very high one of 15 million bps or about 4 times that of commercial television.

It should be noted that facsimile transmission of the type just discussed is a highly inefficient way to convey information. This is so because much of the field to be transmitted is taken up by white area which contains no information, but a high degree of resolution is required to provide sharp black-white transitions.

An interesting comparison can be made of the various ways to transmit the printed page if we assume that a reader needs about 2½ minutes to cover a single-spaced typed page. In order to keep up with the reader we would need a speed of about 220 bps for a teletypewriter, 3,300 bps for facsimile, and 15 million bps for television at the usual frame rate of 30 per second. Obviously this frame rate is very wasteful if we need a change only every few minutes. The bit rate could be greatly reduced by reducing the frame rate or by transmitting only a portion of the page at a time.

Telewriting represents a form of communication that resembles both symbolic and facsimile transmission but does not fit very neatly in either classification. In brief, it provides a means for tracing at the receiving end the lines drawn by the pen (or stylus) of the user at the transmitting end. Thus it is capable of sending handwriting, line sketches, pictures, or any other symbol that can be created with a pen or pencil. There are a number of ways to accomplish this, and the signals required for transmission to a distance can readily be handled in the voice telephone band.

Communication Networks.—A number of users may be interconnected by a network of communication channels so that each may communicate at will with the others. If the number of users is small, each may have a direct connection to each of the others (fig. 44) and this is often referred to as private line or full-period service since the channels are solely and continuously available to the users. If there is a large number of users, the number of interconnections becomes very large (approximately equal to $\frac{n^2}{2}$ when n is large). Also, if the number of users is large, it is unlikely that each will want to talk to all the others simultaneously. In this case the number of channels can be reduced by running a channel from each user into a central point (switching office) where it can be connected (or switched) to the channel associated with another user (fig. 45A). In the case illustrated, by making three connections at the office, three simultaneous communication links can be established, provided the communication needs of the users do not conflict. The switching arrangement of figure 45A greatly reduces the number of channels required, but if there were large distances between users the channel mileage might still be large. For example, if three users were in Chicago and three in New York, there would be three channels from each of these cities into a central point, say, Pittsburgh. The mileage can be reduced by using the trunking principle (fig. 45B). In this case a switching office could be set up in each city and the two offices connected by a channel (or trunk). There is no saving if everyone in one city wants to talk all the time to someone in the other but this is not usually the case. The users sometimes talk to no one and at other times they will be talking to users in their own city. Therefore, it is possible to use far fewer trunks between cities than the number of users in the cities.

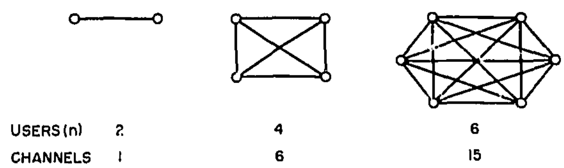


FIGURE 44. *Full period service networks.*

The commercial telephone system is an enormous switched network using hundreds of switching offices all interconnected by trunks so that any one of about 80 million telephones in this country can communicate—sometimes by traversing a half dozen or more offices—with any other telephone here or abroad. In the United States, calls over the nation-spanning network are now commonly established by the dialing of a code by the user and are referred to as DDD (Direct Distance Dialing) calls. Such a network is also referred to as a "common user" system since it is available for all users and for many kinds of services.

Switched common user services are also evolving to handle digital signals such as those required for teletypewriters and slow-speed data. Dial TWX

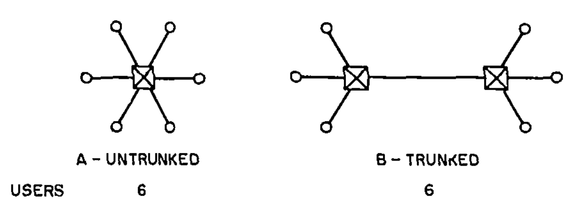


FIGURE 45. *Switched service networks.*

(DTWX) and Telex are examples of services analogous to the voice service of DDD.

It should be noted that switching is of most advantage where a user wishes to have the ability to communicate with many others and does not expect to have a particular connection set up for long periods. There is little advantage if only a few points are involved or where connections need to be made for long periods. Where . can be used efficiently switching brings great economies, but there are penalties too. Signals tend to deteriorate as they travel over a channel; some of this deterioration is proportional to the number of switching points for reasons beyond the scope of this paper. Thus the maximum bit rate is lower over a highly flexible switched network than with a direct connection.

Communication Channels

Potential Channels.—Commercial communication systems are potentially capable of furnishing channels for handling a wide range of information rates even though not all of these are immediately available to users.

The most common channel is that used for voice transmission. This channel can be obtained either on a full-period basis or for short periods on a switched basis. The band used for speech transmission is about 3,000 cycles wide but, as we shall see later, a smaller amount is usable for digital signals.

Telephone channels that go over distances more than 15 to 25 miles are commonly combined into blocks[33] which are transmitted as a unit over a set of conductors (or radio path) thus sharing the cost of the conductors, amplifiers, etc., among the channels. While the number of channels transmitted over a system depends on various circumstances, the total block, regardless of size, is assembled out of basic building blocks or modules that are more or less standard throughout the world. Thus individual channels are assembled into a "group" of 12; 5 of these groups, in turn, are put together to form a "supergroup" of 60 channels, and 10 of the latter form a "mastergroup" of 600 channels. While voice channels are assembled in this particular manner, there is no reason why some other combination, say a block of 2 groups or 5 supergroups, cannot be made available for data transmission. Thus a tremendous variety of bands is potentially available.

Before proceeding with this line of thought, we need to clarify the term "bandwidth" which is used in a number of ways by communication people. In assembling voice channels into larger blocks, some frequency space is unavoidably used up as "guard space" between channels to prevent interchannel interference. It is customary to use 4 kilocycles (kc) total space per voice channel and this is referred to as nominal bandwidth. A 12-channel group will therefore occupy a nominal band of 48 kc. Of the 4-kc nominal band, about 3 kc (300 to 3,300 cps) is usable for speech and the remainder is guard space. However, the filters which provide channel separation influence the channel characteristics well inside the 300 to 3,300-cycle band. The effect is small on voice transmission but fairly large on data transmission. Thus the "usable" bandwidth for data is nearer 2 kc. For bands wider than voice the percent of usable band becomes greater; a 48-kc channel, for example, has about 40 kc of usable band for data. A few of the bands existing in communication systems and their approximate utility for data are:

Nominal bandwidth	Approximate usable band		Approximate bit rate
	Speech	Data	
4 kc (voice)	3 kc	2 kc	2,000 bps (switched). 2,400 bps (private line).
48 kc (12-channel group)		40 kc	40,000 bps.
240 kc (60-channel supergroup)		200 kc	200,000 bps.
4 mc (TV)		4 mc	4,000,000 bps.

Neither the usable band nor the potential bit rate can be stated rigorously. By suitable treatment (equalization) the usable band can be extended but it is usually not economical to push utilization to extremes. Similarly, with suitable terminal gear the bit rate can be made larger than indicated, but this may be a very costly way to obtain more

[33] The technique for combining channels into blocks is referred to as multiplexing and is accomplished by translating the frequencies of the channels so the latter can be stacked one above the other in the frequency scale much as is done with radio or television channels. The technique is also referred to as carrier transmission since frequency translation is accomplished by the use of so-called "carrier" frequencies.

capacity. A rate of one bit per cycle of usable band (as shown above) is usually about as far as it is economical to go at present and often it will prove economical to use rates about half this amount.

A voice channel can also be divided into smaller bands. A common unit is the telegraph channel with a nominal band of about 170 cps, and a voice channel can provide a dozen or more such channels.

An interesting recent development is the use of digital signals for the transmission of speech. This is accomplished by using Pulse Code Modulation (PCM) which typically employs a 7- or 8-digit code to describe the speech wave. When the wave is sampled and the digital descriptions are sent 8,000 times a second, telephonic speech can be faithfully reconstructed. This is a lavish use of frequency space since an 8-digit code requires 64,000 cps to transmit 1 voice channel instead of the 4,000 cps required with analog techniques. However, by regenerating the pulses at frequent intervals it has been possible to use much higher frequencies on conventional wire lines than otherwise possible, and under some circumstances this type of transmission is proving more economical than analog in spite of the wide frequency band required. In a system now going into service, the wire lines operate at the rate of 1.5 million bits per second (megabits per second or mbs) and carry 24 voice channels simultaneously. Thus there is a growing network, presently limited to distances up to about 50 or 100 miles, capable of handling 1.5 mbs (or fractions thereof) with very low error rates because of the frequent regeneration of the signal. Experimental work is going on at rates up to 200 mbs for possible long-haul use.

Available Commercial Channels.—It is apparent that the commercial communication system has grown up in a way that makes it possible to develop channels ranging from the narrow telegraph channel, with nominal band of about 170 cps and a rate of 75 bps, up to the television channel of 4 mc and a rate of close to 4 mbs. Some basic transmission systems will handle bands about double this width and experiments are being made with rates of hundreds of megabits.

However, as noted earlier, not all of these channels are available to the data customer on an off-the-shelf basis. One reason is that not all points in the country have all bandwidths available. Another is that the bands, as they exist in the commercial communication plant, are not directly usable without some terminal equipment (frequently called a data set) to convert the digital signal into a form suitable for transmission over the existing channels. Both of these situations can be dealt with in a straightforward manner and are being taken care of as the demand for wideband data transmission develops.

The voice channel is most readily available for data transmission since it is not only universally available but can be obtained both on a private line and also on a common user switched basis. For this reason, terminal equipment for the use of voice channels has been among the first to be developed. The switched, slow-speed digital services such as DTWX and Telex are also developing rapidly and, in addition, a significant start has been made on supplying high-speed facilities using bands wider than voice (48,000 cps and more).

Figure 46 tabulates some of the typical channels available on an off-the-shelf basis.

Channels may be used in several ways. The most flexible is the Full-Duplex arrangement which provides independent channels in the two directions. Another form is the Half-Duplex arrangement which provides a one-way channel that is reversible in direction. Channels need not be used at the same speed in the two directions. For example, it is possible to have a high-speed transmission channel in one direction with a slow-speed control in the reverse direction.

An auxiliary channel is often used along with a data channel. This may take several forms such as a voice width control channel available along with a 40,800 bps data channel, or it may be a low-speed forward channel to carry synchronizing information for facsimile.

Experimental Channels and the Future.—The development and utilization of more of the potentially available wideband facilities is continuing. There have already been a number of experimental installations of which the following illustrate the wide range of possibilities:

1. An 875,000 bps serial system.
2. A facsimile system using a 240,000 cps nominal band and transmitting 6 pages per minute with about 160 lines per inch resolution.

FIGURE 46.—*Rate of transmission of typical commercially available communication channels*

Type of communication	Private line	Switched voice (DDD service)	Switched digital (DTWX service)	TELPAK*		
				A	C	D
Voice	Speaking rate	Speaking rate				
Teletypewriter	60 or 100 wpm		60 or 100 wpm.			
Serial binary data	75, 150, 1,600, 2,000, 2,400 bps.	200, 1,200, 2,000 bps.	150 bps	40,800 bps		
Parallel 7-bit or 8-bit tape or card transmission.		20 ch/sec, 75 ch/sec.			15,000 ch/sec.	62,500 ch/sec.
Telewriting	Handwriting rate	Handwriting rate.				
Facsimile	2,400 pe/sec	1,800 pe/sec				
Television (closed circuit).	Broadcast and educational TV grades.					

*Bands wider than voice.
wpm=words per minute.
bps=bits per second.
ch/sec=characters per second.
pe/sec=picture elements per second.

3. A private switched system for interconnecting computers at a speed of 15,000 characters per second (105,000 bps).

The pattern in this work has been to devise experimental facilities to meet special user needs and to follow this with off-the-shelf arrangements if the need becomes widespread. This pattern is likely to continue and ultimately the full potential of commercial communciations systems'should be available for data transmission.

Restraints Imposed by Communications Channels.—In view of the large variety of communications channels actually or potentially available it seems unlikely that library communications will be greatly limited by technical matters. Instead, it seems more probable that the choice of communications will be a matter of economics and will reduce to a careful weighing of costs against the services received. We will have more to say about the matter of costs and the selection of economic modes of communication later.

Users may have difficulty in adapting themselves to the fact that communication media are not perfect. Errors of transmission will occur. In most cases the error rate will be far smaller than that introduced by humans. In some cases, however, the accuracy of transmission will not be adequate and techniques for detecting and correcting errors will be needed. These techniques may be very simple or highly complex depending on the need of the user. They are often implemented by the supplier of the business machines but can be provided as part of the communication channel.[34]

Interface Problems.—So far our discussion has been confined mostly to the channels used for communication although some suggestion has been made of the need for auxiliary control circuits and for interconnection between the machine and communication channel. It is now time to examine the complete user-to-user communication system.

A generalized communication system is shown in figure 47. Basically, it consists of customer sending and receiving terminals, a communication set (or terminal), and a transmission line. The last includes, of course, whatever switching mechanism may be required to interconnect customers. The terminals and the communication set are located on the customer's premises, the former furnishing the customer's signal and the latter adapting the signal for transmission over the line. In our previous discussion we have considered the communication channel to be made up of the set plus the line.

[34] This is a highly technical subject which the authors believe is beyond the scope of this paper except to point out the occurrence of errors and the possibility of reducing these to very small rates if it is necessary.

For voice communication the customer merely indicates, through dial operation, the connection he desires and provides the acoustic speech signal. A similar signal is delivered at the receiving end. The communication set is, of course, the telephone set; it provide the control mechanism and converts between the customer's acoustical signal and the transmitted electrical signal. In data transmission, the terminal may be a computer, a console, a teletypewriter, or machinery for sending and receiving punched cards, paper tape, magnetic tape, or graphic material. The communication set is called a data set and not only adapts the customer signal for transmission over the line but also may perform control functions such as selecting the path, automatic answering of calls, notifying the terminal equipment that a connection has been made, and the like. Where auxiliary channels are required for synchronization, etc., the data set derives such channels.

In some cases the interconnection at the interface between the terminal and set is quite straightforward. In voice communication, for example, even though the signal is highly complex and varies greatly from person to person a telephone set has evolved that provides a satisfactory interface for all. This situation has not yet been achieved with data, partly because of the wide variety of data handling machines in existence and partly because the industry has not yet taken full advantage of the possibilities of standardization. Serial binary data represent the simplest form of signal and hence presents the fewest interface problems. Some steps toward standardization have already been taken by the Electronic Industries Association by defining waveform, voltage, functions of the control leads, etc.

In other forms of data transmission the added dimensions of the signal contribute to the complexity of the problem. For example, until a standard code is adopted for tape and similar parallel types of transmission, there is an almost infinite variety of codes and signal characteristics possible. The machine user may easily, therefore, be using a number of digits or other signal characteristics which does not match the capabilities of the readily available communication channels. It is theoretically possible to provide channels which will handle any signal presented but this often results in placing a costly burden on communication.

Many of the interface problems that arise are in connection with analog signals or signals that have some analog aspect. An example of the latter is a pulse train in which the "waveshape" of the pulses is critical. This can occur in some, but not all, systems for transmitting magnetic tape. Another example is the case of high-speed facsimile, where the problem of encoding the signal in a form most suitable for the line may be quite complex. The point to note is that carriers and business machine companies have had considerable experience in the joint solution of these problems and usually they can be handled without undue penalty when they are faced early in the design of a system.

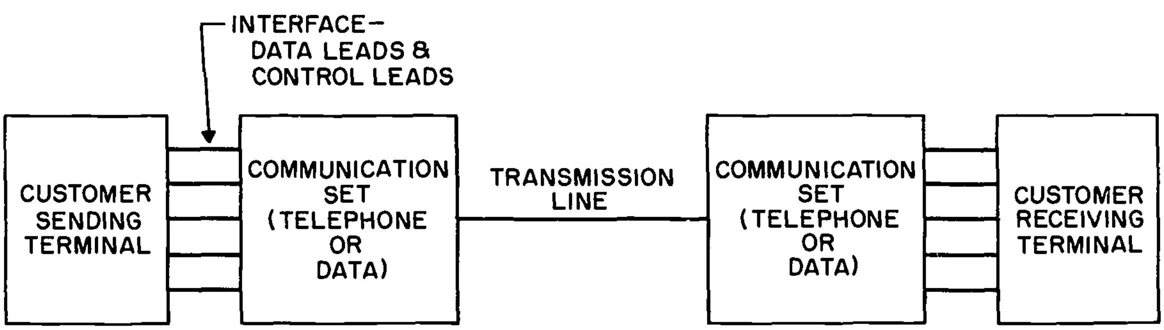


FIGURE 47. *A generalized communication system.*

System Planning

Choosing the System.—In planning the communications for an automated library system one should consider many possible solutions and examine the experience of others who have existing systems with similar operating characteristics.

The electrical communications for an automated library no doubt will encompass a variety of facilities, particularly for remote-inquiry purposes. Consequently, need in function must be clearly determined before a communications system or service can be defined or selected. The function that the system is to perform must, of necessity, be the first consideration. It is defined by answering the following questions:

> What is the communications system to do for the library?
>
> Specifically, what kind of messages is it to move?

In addition to function there are six other considerations:

1. Message distribution
2. Message volume
3. Urgency
4. Message language
5. Accuracy
6. Cost

These are the seven communications criteria. They are used to interpret a business need in communication terms. When this is done, the seven criteria become the "determinants of choice"—the choice of a communications system or service that best fits the need.

Selection and Utilization of Channels.—In selecting communication facilities all seven criteria should be considered; not the least of these is cost, which will be treated in a subsequent section. Function is the first criterion that will be examined. In library communications a variety of functions may be required, for example:

1. Transmission to and from a console used with a digital computer for output and/or input
2. Transmission of standard digital media such as punched cards, paper tape, and magnetic tape
3. Teletypewriting
4. Telewriting
5. Remote copying of graphic material (facsimile)
6. Voice transmission

Other parameters which influence the choice of channels are the following:

1. Distance between users
2. Amount of traffic
3. Distribution of traffic

Distance affects not only the choice of channels but also their utilization. A complex arrangement like a computer console must send a number of different signals to its computer. If it is located close to the computer then it may be desirable to provide a separate wire (channel) for each of the signals that can be sent. In this case there may be dozens of channels per console but each would be used rather inefficiently. If a distance is appreciable, it usually pays to combine (multiplex) all these signals into a single more efficiently used channel. This multiplexing is achieved by adding complexity to the console or communications system which must be balanced against the channel cost. For short distances it is usually economical to use private line point-to-point channels. But as distances become longer the costs of switching must be weighed against the channel savings which can be achieved.

The most important factors in choosing the kind and utilization of channels are the amount, urgency, and distribution of communications traffic. The amount and urgency of traffic determines the communication rate of the channel. Even a large amount of communication can be handled over very slow-speed circuits if there is no need for immediacy. For example, a printing telegraph operating at the low speed of 75 bits per second can send about 300 pages of information in a 24-hour day. However, if it were necessary to send this information in an hour a full voice channel (operating at 1,600 bits per second) would be required.

A channel need not be used exclusively for a single specific purpose. Several types of terminal devices can be used with a single channel, and voice-band channels can be used alternatively for voice and data. Higher speed channels may also be used on an alternative basis. For example, a 48,000-cps channel may be used either for 12 voice channels or to transmit 40,800 bits per second of serial data together with one voice channel. Another possibility is to use the channel for voice during the day and data at night.

Remote Input/Output Devices.—Remote input/output (I/O) devices that could be used with an automated library system cover a broad spectrum both as to operating sophistication and to cost. The proper device for a particular application will depend on the quantity of information it must process, the information format, the operator, and, of course, the cost.

The teletypewriter is widely used as the inquiry device for business information systems. It is relatively inexpensive and its similarity to a regular typewriter makes it easy to operate. Furthermore, it can be used with a variety of communication channels.

If teletypewriters are used as "on premise" input/output devices for the automated library, simple wire circuits would be the most economical way to connect them with the processing and storage equipment.

If the automated library were to be queried by the remote location at relatively infrequent intervals and with short messages, a common user teletypewriter service would prove most economical. The monthly charge is low and messages are charged for individually. In a reverse situation where the automated library would call many and varied locations at infrequent intervals, the total volume might justify the use of a "wide area" service.

In cases where there would be a high community of interest among libraries—for example with a branch library system—the high volume of message flow would probably justify private line teletypewriter service. Should there be much telephone calling between these libraries, it might be desirable to send teletypewriter messages alternatively with voice.

While the teletypewriter would be an effective input/output device for an automated library, it would require a trained attendant to operate it. This attendant not only would have to be skilled in the operation of the devices, but would also have to be trained to transmit information in the proper format for computer inquiry. This would rule out teletypewriters as the console to be used by the public. Instead a specially designed, easily operated console is needed for public use.

The automated library could use a combination of teletypewriters and consoles in the same way the airlines do for their computer reservation system. With such systems, the consoles, called "agent sets," are used at points where the volume of reservations is high enough to warrant this expensive equipment. At the low-volume points teletypewriters are used. However, remote consoles need not be as complex and expensive as the airline-agent sets. Stockbrokers, for example, use an inexpensive device to obtain information about stocks from a central computer record.

One reservation system has another design feature that might be applied to the automated library. From the central computer which stores the reservation information radiate a number of high-speed (2,000 bits per second) communications channels. At strategic points along these channels there are speed-change buffers to which are connected the circuits from agent sets and teletypewriters. By this technique the low-volume, low-speed messages from the remote reservation bureaus can be fed efficiently into the central computer and vice versa.

With the library system, a number of remote libraries could have low-speed communication channels feeding into a regional center. . the regional center there could be high-speed lines feeding directly into the central library computer and speed-change devices which buffer the low- and high-speed channels. A "time-division multiplex" can also be used to connect a number of low-speed channels to a high-speed channel. If any point should have sufficiently large message volumes, it too could have a direct high-speed channel into the central computer.

The principal drawback of keyboard equipment being used as inquiry devices is the frequency of keying errors made by the operators. One way to overcome this problem is to prepare messages "off-line" on a machine capable of producing both hard copy as well as paper tape. The printed page (hard copy) serves to check the accuracy of the information on the tape. After the necessary editing the tape can be sent on to the library computer.

If there is no urgency in the sending of the messages, they can be accumulated on tape for transmission at a later time. This data concentration

735-898 O—64——15

for batch processing can be done at each remote library or at regional centers. By scheduling such delayed transmission until off hours, messages can be sent over channels that are used for voice communications during the working day. The remote equipment can be arranged to transmit the tapes automatically upon command from the central computer. An insurance company uses just such a scheme. All the daily transactions of its district sales offices across the country are transmitted during the night hours into the data center at the company's headquarters. The center polls each office when it is ready to accept its information. In this way the workload of their processing equipment can be balanced efficiently.

Off-hour transmission is also a practical way to send facsimile messages. Present-day facsimile equipment uses bandwidth very inefficiently and may be costly if channels are provided for this purpose alone. Rather than providing additional communications channels to transmit graphic material during working hours, it may be practical to transmit it over a channel or group of channels after working hours when they would otherwise be idle. Coding techniques to increase the efficiency of facsimile transmission are being studied and offer the possibility of a tenfold improvement in the future.

Television is another method of transmitting graphic material but it also uses bandwidth inefficiently. Yet this means of communication is very effective for some applications. In a library, a closed circuit TV system could be used by the public to obtain data from an information center. Soundproof booths equipped with TV monitors and telephones would enable the patrons to contact attendants at a central location. Upon request, information would be displayed on the TV screen, and perhaps quick-print cameras could be used to photograph the display for later reference. Commercial television standards are not good enough to handle a whole page at a time but would probably be quite satisfactory for material the size of a catalog card. For on-premise service the added bandwith required by TV would be of no moment, but for long distance transmission cost consideration would probably dictate the use of facsimile.

Systems have been designed which use recorded voice messages as replies to digital questions addressed to a computer. The advantage of these systems is the simplicity of the terminal equipment. A simple number keyboard and telephone is all that is required at the remote locations.

Low-cost card-reading devices are available to read information stored in punched cards and to convert it into electrical signals. They also have keyboards for generating alphabetic and numeric information manually. The electrical outputs from these devices can be transmitted over regular telephone channels (alternatively with voice if desired).

Telewriting devices can perform functions in some applications that are not possible with other I/O devices. For example, they can be used when validation by personal signature is required to prevent fraudulent authorization. Or they could be used in an automated library system to allow the public to transmit order forms to other libraries perhaps by means of notations in suitably assigned spaces. At the receiving locations these order forms could be read by optical scanners for input into computers.

Voice transmission should not be overlooked in planning the automated library. It can provide a very simple means of obtaining information from attendants at the central processor when other schemes prove too expensive. Here, too, the public could make a permanent record by recording the voice replies.

These, then, are a few of the ways in which electrical communications can be applied in an automated library. Many variations can be devised to meet particular requirements. Only ingenuity and funds will limit what can be done.

Communication Costs

General Considerations.—Perhaps the most difficult task in planning the communications for an information processing system is deciding how much to spend. For optimum system design communications should cost neither too much nor too little but should be in balance with cost and the service objectives for the entire system.

The principal advantage of electrical transmission of data lies in high speed as compared to physical transportation. Electrical communications will prove in whenever the value assigned to the higher speed of information movement can offset the higher cost of this medium. Such value can be translated either in terms of dollars or in terms of service and convenience. Broadly, system costs are determined by four factors—bandwidth, distance, volume of data, and data format—and a careful cost analysis of these factors is an essential part of choosing the right facility for a particular situation. It is difficult to examine these factors individually since they are so closely interrelated. A better way is to consider how they are reflected in modern service offerings.

Electrical communications can be provided by privately owned or leased systems, or they can be obtained from common carriers. Privately owned and leased systems vary greatly in their characteristics and cost depending on the type of facility, reliability, length, complexity, etc. Because of the wide range of costs possible with private systems, we have chosen to simplify our discussion by restricting it to common carrier offerings. Even with this restriction, the range in costs can be fairly large because of the many kinds of service offerings; we have therefore limited ourselves to a few cost examples which illustrate the effects of the factors mentioned above. These examples are not sufficient for guiding system design and, of course, should not be construed as price quotations.

Common Carrier Services.—There are two general classes of communication services available from the common carriers in this country. In one class (private line) a person has the exclusive use of the channels; in the other class (common user) he shares channels in common with others. Today every corner of the nation is linked with both common user and private line networks of various types—telephone, telegraph, television, etc.

The choice between private line and common user telephone or telegraph facilities depends on the volume of transmission and the numbers of locations that must be reached. If the volume is large (or there are large numbers of short messages) and there are only a few relatively fixed locations involved, then the choice probably would be for private line. On the other hand, if the volume and number of messages are small then regular dial common user service should prove more economical.

Recently, a "wide area" service offering was initiated in the dial telephone field. This new service gives the large volume user, who makes calls to many and varied points, an opportunity for significant economies. A similar offering is being planned for teletypewriter and low-speed data.

The person using Wide Area Telephone Service (WATS) is connected to the nationwide telephone network through special access lines and may choose either full time or measured time service. With full time service, calling within a selected territory (zone) is unlimited, within the capacity of the access line. Charges are based on a fixed monthly rate per access line and depend on the zones served. With measured time service a person can generate up to 15 hours of calls per month to the selected zone for a fixed monthly fee. Calls in excess of 15 hours are charged on an hourly basis. Out-of-zone calls can be made over regular lines and charged on the same basis as regular long-distance calls.

Figure 48 shows a map of the United States divided into he wide area zones available to a person in Washington, D.C. The territory of a zone includes the area labeled with the zone number plus the territories of lower numbered zones.

At the present time if a person wants to use a broadband (wider than voice band) channel, he can only lease private line facilities. There are no common user broadband service offerings at the present time; however, this kind of service is under study by the common carriers.

There are two general types of private line offerings: regular leased service and TELPAK service. With regular leased service a person contracts for the exclusive use of a channel facility for a fixed monthly charge. The facility is engineered to meet the particular criteria specified by the user. The charge is based on the bandwidth of the facility, any special conditioning required, and the distance. TELPAK is for the communication user who desires to buy bandwidth in bulk and have it subdivided into channels of specified width to satisfy his individual needs. The conditioning on a channel is applied at an additional charge to the user, but the local loop—the channel from the local central office to the user's premises—is included in the TELPAK charge.

The charges for long-distance telephone services are the same when used for data transmission as when they are used for voice. The only difference is that when the circuits carry data an additional charge is made for special equipment needed to provide the interface between the business machine equipment and the communication facility.

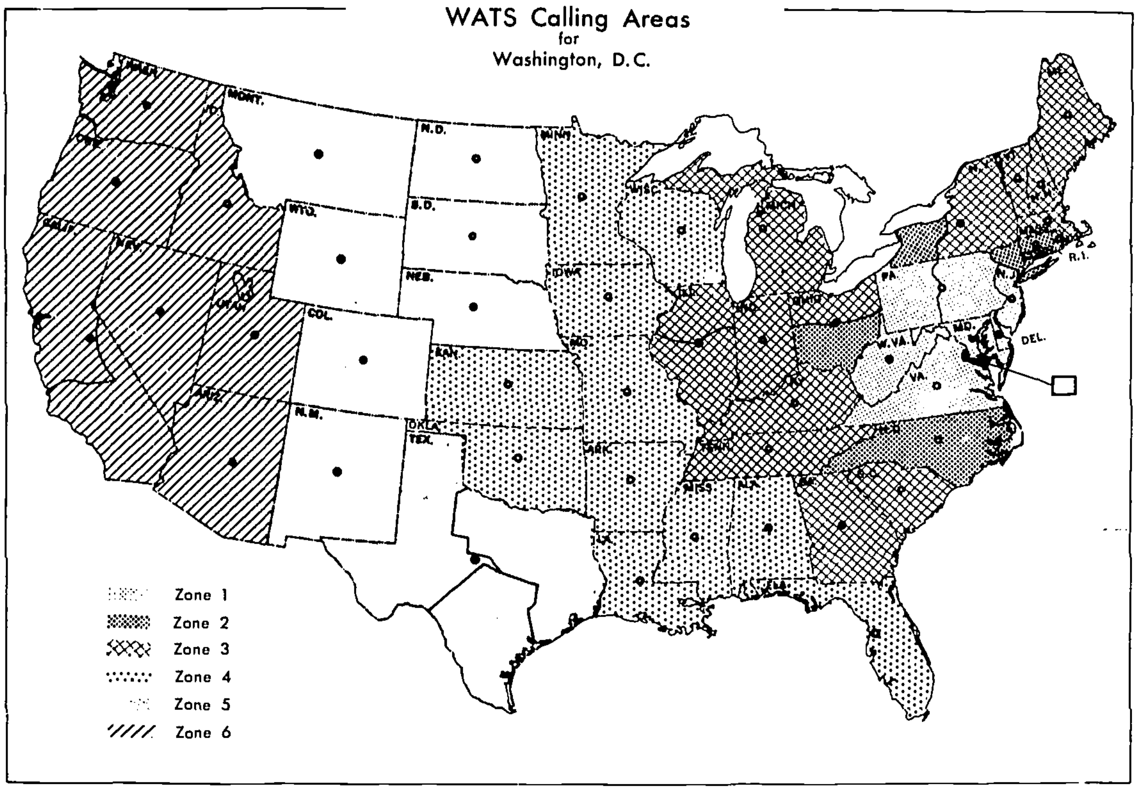


FIGURE 48. WATS *calling areas for Washington, D.C.*

Cost Illustrations.—The material in this section is presented to help the reader understand how communication costs are influenced by bandwidth, distance, leased or shared use, wide area concept, type of terminals, etc. It is included for illustrative purposes only and should not be used as the basis for system price estimates.

The cost of bandwidth can be illustrated with typical TELPAK line charges:

Designation	Bandwidth in terms of equivalent voice channels	Cost in dollars per mile per month
TELPAK A	12	15
TELPAK B	24	20
TELPAK C	60	25
TELPAK D	240	45

It will be noted that the cost per unit bandwidth becomes much lower when it becomes possible to use large blocks.

For very long systems the line costs represent a large part of the total, but for short systems (and some data formats) the costs of terminals may be significant. This is illustrated by figure 49 which shows the cost for a 1,050-word-per-minute tape-to-tape data transmission system as a function of length, using two types of facilities:

1. Voice private line
2. Voice channel in TELPAK A (75 percent fill)

The lease charges for the tape terminals and data sets are common to both arrangements. However, with TELPAK there are additional termination charges. All of these are fixed monthly charges independent of distance.

The curves show that for distances of a few hundred miles the fixed terminal charges may be

50 percent or more of the total charges. Obviously, these terminal charges are highly dependent on the type of data equipment used; paper-tape terminals are shown in the figure merely as an example. The figure also shows minor differences between the private line and TELPAK costs, but these comparisons obviously depend on how efficiently the TELPAK A channel is put to use and how costs are prorated among the services handled. In the example given, it has been assumed that 75 percent of the TELPAK channel will be used (i.e. 9 voice channels for all services) and, therefore, the pro rata cost of the one channel used for data is one-ninth of the total TELPAK cost.

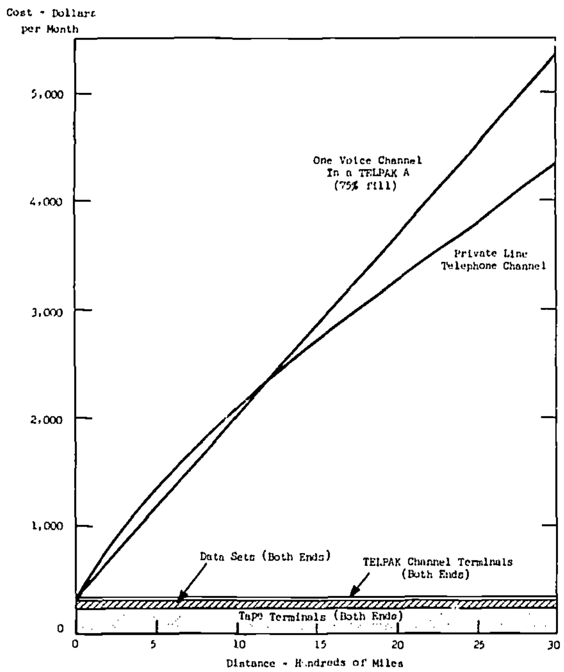


FIGURE 49. *Data transmission—cost vs. distance, using a voice-grade channel with a 1,050 word/minute paper-tape-to-tape terminal.*

So far we have been discussing costs of private line service. It is interesting to compare this with common user service. To do this, it is necessary to take into account the volume of data per day and the length of the transmission period. This is brought out in figure 50 which shows the communication cost to transmit, for 1,000 miles, varying amounts of information at 1,200 bits per second over 4 types of services (costs are for the transmission facilities only):

1. Private line telephone channel
2. Wide area telephone (full time WATS)
3. Voice channel on 75 percent loaded TELPAK A
4. Regular long distance calls (DDD)
   a. 1-minute messages
   b. 3-minute messages

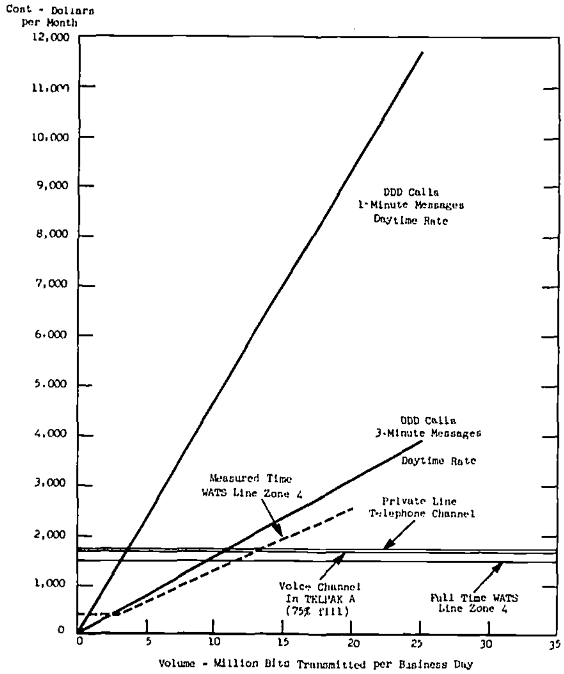


FIGURE 50. *Data transmission—cost vs. volume using a 1,000-mile voice-grade channel—1,200 bits/sec.*

The costs for private line, WATS and TELPAK are fixed and not dependent on usage, while the DDD costs are based on the number of calls and length of calls. In the Bell System, there is a minimum charge on the telephone message network for a message of 3 minutes or less. Therefore, the minimum charge per bit is obtained when messages can be accumulated so that each connection is used for 3 minutes or more. However, if the user has messages lasting only 1 minute and for some reason wants to transmit them immediately (i.e. not waiting until 3 or more have been accumulated), the cost per bit will be 3 times the cost of accumulated messages. The curves also bring out the fact that the telephone message network provides the lowest cost per bit for small volumes of communication (in this particular case for under 5 or 10 million bits per day), but private lines provide much lower costs per bit for large volumes of data.

The very low cost per bit that can be obtained with high-volume usage of wideband private lines is brought out by figure 51 which shows how many bits of information can be transmitted per dollar of channel charges for a 100-mile circuit. Perhaps a better feeling for the low cost of high-volume data transmission is given by the last column which shows that it would cost $60 to transmit *The Rise and Fall of the Third Reich* by 100-speed teletypewriter while it could be sent over a TELPAK D channel for only 55 cents. The assumption in each case is that there is enough need for communication to keep the circuits busy 8 hours per working day.

FIGURE 51.—*Cost of Data Communication*

CHANNELS AND TERMINATING EQUIPMENT*

Type of communication	Million bits per dollar	Cost to send *Rise and Fall of the Third Reich*
100-Speed Teletypewriter	0. 4	$60. 00
DATA-PHONE service (Data Set 201 type) (2,000 bits/ sec)	2. 1	11. 00
TELPAK A (40,800 bits/ sec)	15. 4	1. 50
TELPAK C (105,000 bits/sec)	22. 0	1. 10
TELPAK D (437,500 bits/sec)	44. 0	0. 55

Assumes 100-mile circuit used 8 hours per day, 22 days per month.

*Not including teletypewriter or business machine.

The price to transmit a page of written or graphic material by facsimile equipment varies, depending on the reproduction rate. This in turn is a function of channel bandwidth and sophistication of terminal equipment. The costs of fac-

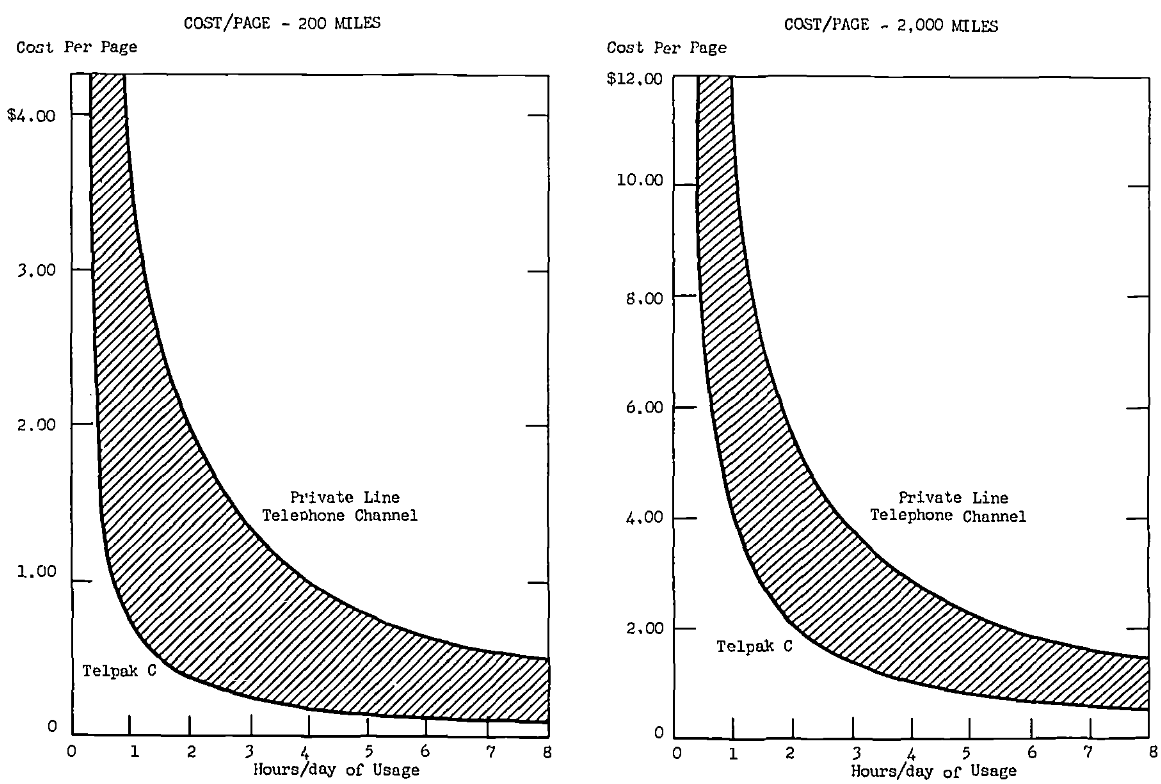


FIGURE 52. ***Cost range of facsimile transmission including line, terminals, and facsimile machines.***

simile per page transmitted are shown in figure 52 for distances of 200 and 2,000 miles. This figure also illustrates the relationship of cost to bandwidth and channel usage. The shaded area indicates the price range: from simple equipment on a private line voice channel to sophisticated equipment using a full TELPAK C channel. The shapes of the curves indicate the lowering in cost with increased use.

It should be evident from this brief coverage that the costs for communications which might be applied in an automated library system can vary widely depending on design objectives. There is no way to estimate these costs accurately in advance of a system plan. Therefore, it would be foolhardy to budget a set amount of money for communications without a system study. Since communications are truly the vital link in an information handling system, they should be planned with the same concern as the other system elements—for optimum performance at optimum costs.

Conclusions

Library mechanization can undoubtedly utilize a variety of communication facilities performing a number of functions. Because of the wide range of communication channels and services either currently or potentially available it does not appear that communication technology needs to restrict the evolution of a mechanized library system. Instead it seems more likely that the degree to which communication over distance will be employed will be a matter of balancing the benefits of geographical extension against costs.

The prudent designer of a library system will do well to consider his communication problems at an early stage to avoid the imposition of unbearable economic burdens by placing unnecessary requirements on the communication system. Some pitfalls to avoid are the use of custom design when off-the-shelf gear will do, the use of real-time transmission when delay will suffice, the use of inefficient modes of communication such as television when motion is not a factor.

It is of basic importance to consider the system—the library, computer, console, displays, and communications—as an entity, not as a collection of parts. The choice of any one of these components of the system is a complex matter which interacts with the others. It will take close cooperation from the very beginning among experts in the fields of library science, computers, graphics, and communications if a sound system is to result. It is indeed a good omen that this symposium has recognized so early the many facets of the problem.

CONFERENCE SESSION VI

Communication in Libraries

HENRY J. DUBESTER

Library of Congress

In these preliminary remarks about communications and library systems, I would like to comment on the communication problem as exhibited so far in the discussion at this conference between the librarians and the technical experts. We have heard the technical people say: "You must state your requirements for us." The librarians say: "Please tell us what you can do for us." This sort of conversation becomes very frustrating; what is always necessary is a progressive dialogue. In my experience with the survey team that worked in the Library of Congress, we had such a dialogue. Progressively we came to know and understand each other more and more. I think that we have come so far that, in addition to the so-called library type and the so-called computer type, we now have people, some in this group here, who are neither the one nor the other, but both.

Ultimately we must recognize that automation will stand or fall, and here perhaps I am reflecting a personal view, on the question of costs. Now the criteria can become very illusive. They may be social judgments as to what is valuable, rather than just immediate savings determined by comparison with the cost under our present methods. It may very well be that some of the costs of research, development, and implementation will be greater than any one library can bear. It may well be that the costs will have to be shared by the library community. As a matter of fact, the needs of libraries at the present time are not the needs of a given library; they are, in my opinion at least, the needs of the library community. In a sense, therefore, when we begin to speak of the library system, we must have a broader horizon and must consider the system of libraries, which we in part represent.

These libraries are not facing a *future* communications network. They have always had a communications network. Communications between libraries are nothing new. The very existence of libraries themselves is posited on the need for communication, and libraries have found that they can extend their resources by cooperation, which is relatively equatable here with the communication process.

Let me give you some examples of such communications. When libraries purchase LC cards, communication is taking place. When libraries make inquiries to the National Union Catalog, or to regional union catalogs, or to each other, they are communicating. When libraries refer reference requests which they cannot handle to other libraries having special collections, they are communicating. Communication becomes very intense on any university campus where departmental libraries have decentralized. With automation we face the prospect that much of what we are doing will be intensified. Communication will increase, and patterns of communication will undoubtedly change.

As an example of library communications take the case of the National Union Catalog. At the present time, most inquiries are received by mail. Only 3 percent of the inquiries received by NUC in a 3-month period utilized TWX, a relatively fast mode of communication. There is a permissible inference on a low evaluation of library service or upon the need for speed in such service. I am frequently tempted to conclude that this evaluation is perhaps attributable to librarians even more so than to the public which does not exploit libraries to the degree that it should. Certainly, until libraries are valued in the manner that their potential justifies, they cannot secure the support which will undoubtedly be required for successful automation, in which more rapid and more sophisticated communication modes will play an essential role.

Communication Systems for Libraries: Some Examples and Problems

J. W. EMLING
Bell Telephone Laboratories, Inc.

Communications in the System Design

It is not my intention to summarize the paper. I should like, however, to review the conclusions, which are roughly the following: It seems very obvious that library mechanization can use to advantage a considerable variety of electrical communication facilities. As we looked over the library field, it did not appear to us that technology would really limit the application of communications in this field. The thing that could very well limit this is the matter of economics. Communications may cost a little or they may cost a lot—it depends a great deal on how they are used. And it will, of course, be important not only to use them correctly but to examine each situation to see that you are getting your money's worth. We are not trying to sell communications to anyone unless the system is really needed.

I should like to emphasize that because a communications system can be either costly or very cheap depending on how it is used, it is extremely important that the designer of a library system begin to think at the very outset about how he will use communications. It would be very unfortunate, indeed, if you designed a system and when you were through you brought in the communications people and said, "Here is what we want to do," because it is perfectly possible that you have put some economic burden on your system which should not be there. There are many pitfalls in the use of communication. You can, just as in the use of computers, without realizing it call for the use of a custom design when something off-the-shelf will do the job. You can ask for real-time transmission when some delay will suffice or for inefficient modes of transmission. Television when you do not need motion is an example of this kind of thing. The only thing that I feel that we ought to emphasize is that as you design a system keep in mind that communications are an integral part of the system and should be considered from the very beginning. If you do this, you may well find that you can communicate over a distance in a manner that will not place any unnecessary burdens on the system.

Cost Examples

I feel that I have an apology to make, because the title of our paper, "Library Communications," is misleading. I confess that while it deals with communications it has very little to do with libraries. Unfortunately this seemed to be necessary because, first, we do very little about libraries. (We are very knowledgeable now after listening to you people for a few days.) Secondly, it seemed very difficult to find out just what an automated library system was. Now we still have this problem to some extent because I am not quite sure that the automated library system has been spelled out in detail as yet. Therefore, it was a little bit difficult, for example, to come up with cost figures other than the general ones we have presented in order to illustrate some of the important considerations in selecting communications. Dr. King was good enough to write, after we submitted the paper to him, and suggest that we pick a few examples to discuss that would be more particularly related to the library field than the ones discussed in the paper. He also suggested some examples, for which we have worked up some rough cost figures. I will show them to you now.

The first example (fig. 53) assumes a user at a remote console with a cathode-ray-tube display. He has a magnetic-disk buffer which holds 1,000 rasters to store non-moving material. In other words, this is a cathode-ray-scope display, but it is not television; we are not introducing motion,

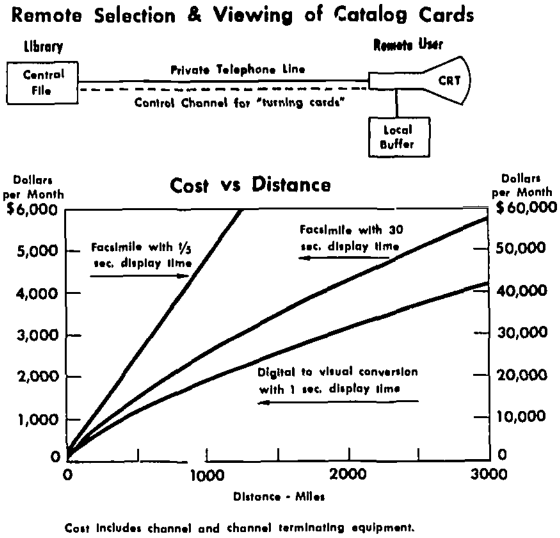


FIGURE 53. *Cost, distance, and transmission modes as factors in remote card file selecting and viewing.*

we are storing information for you. Dr. King says the user can control remotely the turning of the cards in the central file. He may want to study each card for various lengths of time and then refer back and forth to material in his local buffer. When the user wants intermittent access to new cards, with display in less than ⅕ of a second, what communication systems are recommended when he is 100 feet away, 100 miles away, and 2,000 miles away?

Before going into this, let me make a comment on the problem. First, let us dismiss the 100 feet away, because if your communications are for 100 feet we can neglect them. You run, for a nominal sum, the necessary number of wires, and it just does not enter into the problem in comparison with the costs of consoles and everything else. But you will notice that this is the console with the dialogue approach that was discussed at length yesterday. I should say something about this ⅕-second response; we assumed that meant that after a request has been turned in to the catalog the card would be displayed in ⅕ of a second. This has quite an important bearing on the problem, and I might note, just by way of passing, that this is indeed quite a short time. If we look some years ahead to the time when the Bibliothèque Nationale might be connected into this system with communication by way of a satellite system, we would essentially use up this ⅕ of a second merely in traveling there and back utilizing the speed of light. So this is indeed a very short period of time. As you will see, we thought maybe people could adapt themselves to something just a little bit longer.

I think that maybe now we might take a look at the first slide (fig. 53). At the top we have a picture of what we are doing. On the left is a central file; we have a private telephone line running to the right, which is the console; in there we have a buffer to store the data so that we can accumulate data and throw it on the scope all at once. First we consider graphic transmission, and if we have graphics at the sending end and we want to transmit this display in ⅕ of a second the top curve applies. The curves all show at the bottom the distance running from something negligible up to 3,000 miles on the right. On the left is a scale of costs. There is also a scale of costs on the right, and this is the one we should look at for the ⅕-second display. You will notice that it goes off scale somewhere at about 1,000 miles for a mere sum of $60,000 a month, and frankly I do not think we are going to sell many libraries on this. If, however, you are willing to take graphic transmission and if you are willing to wait 30 seconds for that display, you have the second curve, and for this we use the scale on the left, which is reduced by a factor of 10 from that on the right. You will notice that even so the curve is lower and it goes off the scale at about 3,000 miles at about $6,000 a month.

If you will take digital transmission rather than graphic transmission, then you can use the bottom curve which is still somewhat better. You can go to 3,000 miles for a mere $4,000 a month. I am not sure that we are going to sell the west coast on this or that we should, because there are other ways of doing this for less money. But the important point to notice is the difference that the response time makes in the transmission cost. If you insist on your ⅕ second, I suspect you are priced out of the market even for 500 miles. If, however, you will take a longer response time with graphic transmission, you can get in the market at 500 miles for $1,200 or $1,500.

In connection with the digital transmission, I should have pointed out that not only is this somewhat cheaper than the middle curve which is graphic, but you do get a 1-second response time with this. This is a measure of the difference between transmitting your information as it is on your tape, the basic information in digital form, instead of using graphic information which contains such a very high degree of redundancy. I think it was a very good suggestion to ask for costs on this because it brings out some of the considerations which the designers of the system must keep in mind. If you are willing to do some kinds of transmission in connection with other services, you may be able to combine your efforts and get lower rates.

The next example concerns a presumed daily output of 250 completely edited LC cards done at a console at the Library of Congress. The edited material is on tape to operate a printer; a remote location has a tape-driven printer. What communication link is recommended? At the upper left of figure 54 we have the console in the Library. It punches paper tape, which feeds through an automatic typewriter over a TWX circuit at 100 words a minute into a remote automatic typewriter that gives you hard copy of the library cards. Now again we have a cost-vs.-distance curve; it goes out to 3,000 miles. The left-hand scale shows dollars per month for handling these 250 cards every day. In this case it does not pay to put in a private line as we assumed in the previous example. This can be transmitted over the TWX system, and the costs reflect this. The steps in the costs merely show the finite steps in the tariff for this kind of service. At the bottom of the curve there is the cost of the teletypewriter equipment, just to give you a feel for the cost of the terminal, and above that are the costs of the message charges. Now I don't know why anyone should do this, as compared to dropping these in the mail and sending them air mail for 8 cents, but if you want it, here is what you can do.

You can actually do something better than this if you wish. You can speed up the transmission and send this over a regular telephone call. You can transmit it at about 1,000 bits a second. You put a tape receiver at the receiving end so that you can play that into a slow-speed teletypewriter because obviously the high-speed teletypewriters are expensive, and unless you are just hanging there waiting to see what that next card is going

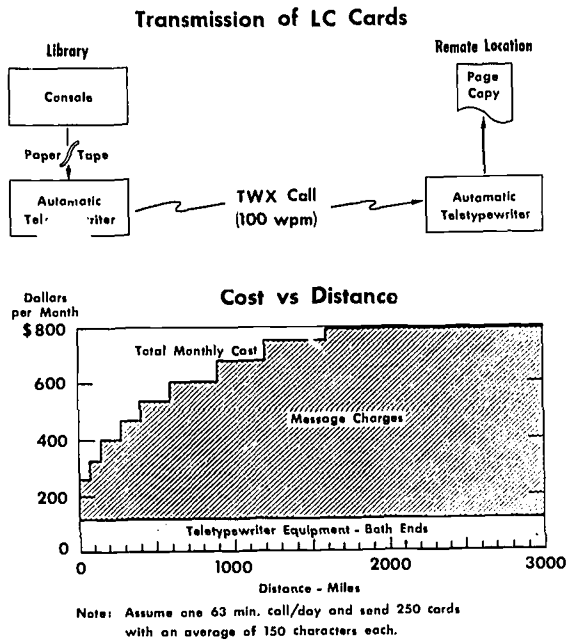


FIGURE 54. *Cost vs. distance in teletypewriter transmission of LC catalog cards.*

to be, a slow-speed is adequate. Now if you do that, that adds a couple hundred dollars down at the bottom for the terminal charges, but it cuts the instrumental cost for the circuit about one-eighth. In other words, that top figure is about $100 per month on top of the $300 or something like that for terminals.

While no one may want to receive his library cards so quickly that he will want to put this kind of service in, the example is extremely useful in illustrating the factors that must be considered and the large range in costs you can get depending upon the requirements you set up. If these requirements are important, you will be willing to pay the cost, but if the requirements are of no importance to you, you can take some delay. You may be able to send these at night, maybe you do not want them to come in instantaneously. You can get a 3-minute call to the coast after 9 p.m. for a dollar; this cuts costs still further. Maybe for some other reason you have a private line to the west coast. If the private line is not used at night and if you happen to have the teletypewriter there and it is not used at night, the answer is obvious; it costs you nothing. So again, if you cut your garment to fit the cloth, you can often get a very nice garment indeed.

Let me take a minute for figure 55. A user at a remote location has identified a pamphlet and would like to have its complete contents in hard copy now. Again, the user is at different distances away and again this is a matter of graphics transmission. Here is an outline of what we do: We have a facsimile transmission system feeding into a Data-Phone over a long distance call and through a Data-Phone and into a facsimile receiver at the other end. It may go over private line instead of the switched telephone plant; your choice depends upon how much you use it. If you do not use it very often, you use the telephone setup. *If you use it frequently, or you have other* uses for your private line, why, obviously this is the way to do it.

The costs of these two lines are illustrated in figure 56. We have here a little different system than on the previous one; the scale at the bottom is the usage measured in number of pages, 8½ by 11 inches, which are set in a day. We have 3 distances, the dotted curves apply for 200 miles, and solid curves apply for 2,000 miles. You have a flat curve, which applies if you rent a private line and use it only for this purpose. For the short

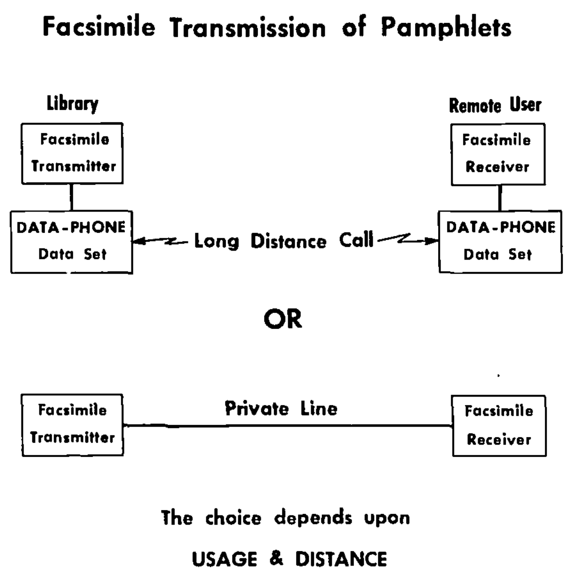


FIGURE 55. *Facsimile transmission of pamphlets.*

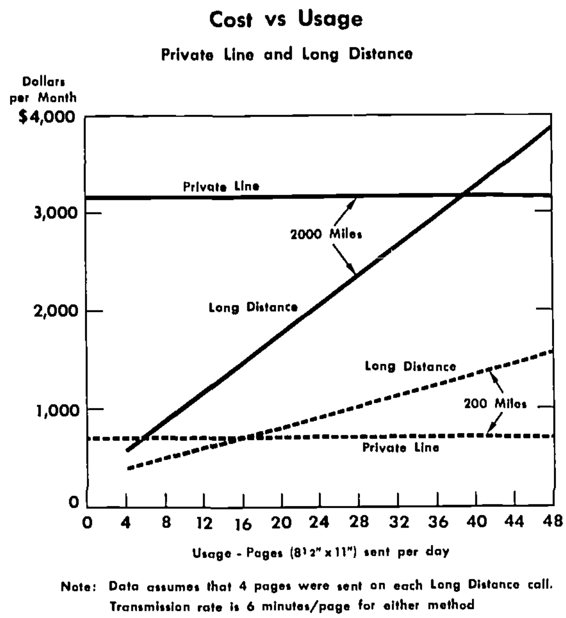


FIGURE 56. *Cost vs. usage (private line and long distance) for facsimile transmission.*

distance, if you get above 16 or 20 pages a day, this is the way to do it, because from there on the costs of your lines do not increase up to the full capacity of the line in whatever period you are willing to use it during the day. For the longer distance it turns out that the break-even point is about 40 pages a day. The costs are somewhere around $800 for this private line for the 200 miles. If you want to go 2,000 miles, it's on the order of $3,000.

The absolute dollar figures are not the things that I would like to emphasize. Sooner or later in designing a system you are going to look at them carefully, and you are going to decide what you can afford. What I want to emphasize today is the wide range in costs that is possible, depending upon the requirements you place on your system; I urge that in designing a system you make your requirments as realistic, or maybe I should say as tolerant, as you can. If you do this and if you can combine this with other services, you may be able to get forms of communication that mean quite a lot to you for a very low cost. If you insist on unrealistic requirements, you can price yourself out of the market.

General Discussion

WOOSTER: I would like to hear more about the possibility of TV on either 3 kc. or 4 kc., without having to lease lines—the possibilities of using regular dial-switch telephone over normal voice channels of real slow-scan TV, say 2½ minutes a page.

I would also like to have one other point discussed. If the Library of Congress builds a central computer to which I can get access over dial-switched-line or TWX lines, can I actually get into the Library of Congress computers by using Telex lines as well as TWX lines without the thought of violating certain consent decrees?

HARRIS: I'll try the one on slow-scan TV. We have a Data-Phone data set for switch connections which I think is along the line you're interested in. I understand that one of the companies that is concerned with terminals for slow-scan television is experimenting with some models of this particular data set and is sending slow-scan signals over dialed-up connections. These are fairly close to a library card I believe, 300 lines of scanning with 230 picture elements along each line. This might correspond to something like 3 by 5 inches with conventional facsimile definition and this goes in 40 seconds over a dialed-up telephone.

QUIRK: I think the question Wooster asked is an important one. I would like to sketch the Data-Phone concept and show how this will tie in. There is a data set at either end of a connection which will take signals from a business machine; in this case let us say that it is a Library of Congress computer. Once past the interface, what happens there is the customer's business. I can well visualize a Telex line coming in here. This means that, in effect, the computer could act as a switching device, take information in and out in both directions acceptably.

ORNE: For something like 4 years we've had a TWX in my library which we use essentially for interlibrary loan. To tell you the kind of bollix you can get into, which is not an engineer's responsibility, about 2 months ago a message came in from Washington; somebody wanted a message conveyed to the chancellor of the university. The clerk who was on duty typed out a return message: "This machine serves only for interlibrary loan."

BRISTOL: Mr. Emling, you mentioned 3 requirements King asked for, one was 100 feet, and one was 100 miles, and the third was 2,000 miles; you discounted the 100 feet as being a negligible factor in wiring. What about a campus with substations within 6 or 7 blocks of the central station? Is this a negligible factor also?

EMLING: Technically this would be handled differently from long distance. It probably, depending on the speed you're talking about, would be handled on the local carriers that we use for telephone conversations; you can handle quite wide bands for short distances on this. It would be a special arrangement; what the costs would be I don't know, but you could find out.

QUIRK: A rough figure for a voice line would be about $3 a mile.

PATRICK: Give me terminal set cost with that.

QUIRK: All right. Again the terminal cost is a function of what you want to get out of the set. This can vary from as low as $5 for a card-transmission system up to several hundred dollars for a more sophisticated type. In this case the terminal gear may be the limiting factor, not the mileage charge.

LIBBY: These are costs per month, right?

QUIRK: Yes.

SWANSON: If I have a telephone line with presumably 3,000 cycles bandwidth all used up for voice transmission, is it my own business what I do with it? Should I choose to spurt a lot of digital data through using up the whole 3,000 cycles, or, if I build something that multiplexes it in such a way as to use up a whole bandwidth, do I pay any more for that telephone line than I would normally?

QUIRK: Again this depends on whether you're talking about dial service or private line service. With a private line service, which you have the full use of, you can do anything you wish within the capability and the characteristic of the equipment. If you want to send more than 3,000 cycles, you're going to have degradation and so on, but this is up to you. But in the dial service there are certain restricted frequencies; if you send on those they would be taken out, because they are used for signaling purposes and so on. We would say that this would be at your own risk. Normally, we would provide the data set because we would compensate. You cannot compensate now for that frequency because we might want to use another frequency later on. In other words we couldn't dedicate a frequency to your use, but by getting the data set and the dial service from the telephone company, as the technology changes, the data set would be changed at no cost to you.

EMLING: There is one other consideration. There are limitations on what you can put over a telephone line. If you put over a telephone line an amount of power greater than the average power in a voice, it will interfere with other voice circuits and may cause serious trouble to an entire system. Obviously we design these circuits for the voice, and we don't build in a lot of extra margin just in case someone comes along with this other need, because this would be a very wasteful thing to do from an engineering viewpoint. Whenever we use a circuit for some use other than voice, even a private line circuit, we are obliged to put on some restrictions as to the level or amount of power you can put over it. We will ordinarily provide a coupling unit that will see to it that you don't exceed this.

SWANSON: That shouldn't cause any problem though, should it? You could always cut the power down to whatever maximum you stipulate.

EMLING: You can if it will do your job. Sometimes it's a neat trick to get your job done if you cut the power down too much.

QUIRK: Another point is that most business machine companies providing this equipment for various users work very closely with the Bell Laboratories in setting designs that will match these requirements.

FUSSLER: May I ask a question relating to one of the curves in figure 56 which you said was the number of pages of facsimile transmission. You indicated that the curve did not show the limits of the full capacity of the line. What is the capacity? The scale went up to 48 pages, I think.

QUIRK: That was a 6-minute transmission, $1/10$ of an hour per each page, so the limit would be the number of hours per day divided by $1/10$ of an hour.

BOWLING: I noticed that in developing your costs you included the tape-producing device, the hard-copy device, and the transmission system, but you didn't include any means of error checking and retransmission; I'd like to know what this adds to the cost. And a second question: When you are using the regular dial system between any two Bell telephone sets, you often go through different cities between two remote points. What is the percentage of time that you get a line suitable for digital transmission between two cities, such as Washington, D.C., and Los Angeles, California?

QUIRK: Let me take the first question. In the error-checking question you have to define what kind of information you are sending. If it is page copy, there's a tremendous amount of redundancy in the copy itself, and an error may not be critical.

BOWLING: I refer strictly to the digital transmission at 1,000 bits per second.

QUIRK: Well again, you could be sending page copy at a 1,000 bits per second. All you've done by using the 1,000 bits is speed up the transmission and cut down the transmission time.

BOWLING: It's the data, the Baudot code, where you are transmitting alphanumeric information.

QUIRK: If each character is important, then there is need for error checking. This can be done in a number of ways; the present one that was costed out in these figures was the so-called Data-Speed. There is also the Data-Speed 2, which we don't have in these figures; the Data-Speed 4 is coming out within the year and will have error detection and block retransmission; the price will probably be somewhere around 150 percent of the terminal costs calculated here.

DUBESTER: Can you translate the problem and the answer for the nontechnician?

QUIRK: The problem of error means that either a piece of information was sent, let's say because of electrical storm disturbance, which was not supposed to be sent, or information was deleted for some reason so that the number, the shape, or the information that is to make up an alpha or numeric character is erroneous. There are a number of ways of checking whether this has happened, but one of the simplest is the so-called parity method, by which you count the number of marks. These marks are added up and, if they're even parity, you check this and put in the proper additional digit to make the even or the odd parity. So that when you count the number of ones or zeros you have received, if you find there are a certain number and the parity does not check, you know there is an error.

Block retransmission means that you wait until you get a series of characters, perhaps 20 to 30, and check the whole group for an error. If you detect an error, then you contact the sending end and request that the whole block be sent again. This is what is called block retransmission.

The second question concerned the dial service. In dialing a telephone call the number of possible paths that you can take through the vast telephone network is unpredicted. No engineer in the Bell System could tell you what particular path a call is going to take. The question was: When you have this data service what assurance do you have of your getting 100 percent accuracy? Again, we don't guarantee 100 percent; we say that any dialed-up connection should get your data through, but there will be an average error rate; we can't guarantee it, but it's on the average, one error in 10^5 with some sort of normal distribution.

UNIDENTIFIED: Last winter Xerox demonstrated a campus facsimile-transmission system in Rochester, I believe. Did that go over telephone wires? Do you know anything about it?

EMLING: We did not provide the lines for that system, but we are either just now starting, or will very soon start, a rather large trial of a similar system for a large company, connecting a number of their plants. This will go over the Bell System and will be used for sending all kinds of information where they have to get the original copy, for example, waybills and orders. It will prevent the necessity for copying, and it will show the nature and appearance of the original. This is one of the best ways, of course, of getting error detection because your eye will tell you pretty quickly whether you have good copy or not. It's an expensive way to do it unless you use a lot of it. This is the moral of the whole business: communications in single packages can be expensive; in bulk, quite inexpensive.

CLAPP: In the library business, we are very dependent upon intercommunications. In 1950 Louis N. Ridenour pointed out that the book funds of 26 principal libraries in the country at that time amounted to a sum in excess of 5 million dollars.[35] He pointed out that this sum would support a pretty fancy communication system and still leave a sum of money for acquisitions larger than was at that time available to any one of these 26 libraries. Nobody has taken up that challenge. We haven't worked it out, and as a matter of fact each of those 26 libraries has worked to get more book funds to buy more books to have locally. What this means, I think, is that we would much rather have a book which we can put our hands on, even though it's in a distant part of the stacks, than be dependent on getting it from somewhere else no matter how good the communication system.

I would estimate that 80 percent of the books that you don't have locally and that are wanted, not by undergraduates, but by the faculty and the graduate students, you can get via interlibrary loan. Now you can't get them overnight, but, if you could get them in 2 or 3 days, you'd be happy. But you can't get them in 2 or 3 days; it takes that long just to do your paperwork, and there are 2 or 3 days of paperwork at the lending end. This boils down to approximately a week, and I think a week is probably optimistic.

Now what Emling and his colleagues offer us is the possibility of reducing this transmission time from days to minutes. What is it worth to us? Do we take any advantage of it at all? How much would it cost, Mr. Emling, to send me—2,000 miles from Washington—an LC catalog card in 1 minute from the time that my order is received?

EMLING: Your library card you can get over TWX in a minute; in fact, you can get four of them in a minute.

CLAPP: How much will it cost?

EMLING: Oh, for 2,000 miles it might cost you somewhere under $1.75 or something like that. But this is sort of inefficient; that $1.75 buys 3 minutes, so they're even cheaper by the dozen. You can get a dozen for $1.75.

CLAPP: I ask the assembled librarians here, is it worth it to pay a $1.75 to get a dozen catalog cards in minutes, or even in the same day, as opposed to getting them a week or 10 days later?

BLASINGAME: Any State library which does any volume of interlibrary loan at all can make the average research library look awfully sick on this issue, because we get requests in and get them in the mail while you guys are still looking for the ball of string. We are set up for this. These are roughly the statistics. In Pennsylvania, with a rather poorly developed interlibrary loan, partly

[35] Ridenour, Louis N. *Bibliography in an age of science. In* Bibliography in an age of science. Urbana, University of Illinois Press, 1951. p. 5–35.

because of the weak libraries we have in the State, we will lend about 30,000 items a year, of which 20,000 will go into the mail the day we receive the request. About two-thirds of those are subject requests and not author-title requests. In other words, we usually have the material in the mail the day the request comes to us regardless of the type of question.

CLAPP: What's the elapsed time from inquiry to receipt?

BLASINGAME: We are trying to get at this problem because we think that we are pretty slow, too. Of the one-third which we don't satisfy, we send about half on to the Union Library Catalog in Philadelphia. We get back locations on about 90 percent of these; in other words in terms of satisfying the requests, I would say we do reasonably well and in terms of the time, I would say again that we have our stuff going while others are still looking for something to wrap the book in.

What I'm concerned about now is how to use equipment available today to speed up this process. In particular, why don't we pay the telephone bill for the requesting library to call us? In short, I think we could do some things with existing systems and without spending any particular amount of money in terms of investment by just raising the operating costs a bit to cut down on the total elapsed time. Second, I think we are getting to the point where we can send the material back to the requesting library on channels other than the U.S. mail. We have buses running around Pennsylvania that get there a lot quicker than the U.S. mail, and for very nominal fees they'll carry small packages. I would like to see someone attack this problem of sending material as quickly as possible without any significant investment cost.

DUBESTER: It may well be that in the confines of a State 90 percent of the requests can be met through a combination of immediate response and a union regional catalog service. There may still be a 10 percent increment which, if it is needed, can warrant more sophisticated applications. These are the questions which I think the library community must also ask itself.

SPARKS: Suppose a catalog card is in a punched-tape equivalent, a form which is equivalent to punched tape, of about 400 characters average. What's the cost of transmitting this, so that I can make, in my own library, a set of cards from this tape?

EMLING: I don't feel that I can give these offhand answers, but there's no reason why I can't send punched tape just as well over TWX as anything else. Twx is 100 words a minute, or about 100 and some bits a second, so you can very easily figure this out, depending on what kind of a code you want; but say 100 words a minute with your 400 characters divided by 6, that's about 70 words, so that it's less than a minute. You can get it in the 3 minutes over TWX or if you send it at 1,050-bit Data-Speed over a telephone line you get roughly 10 times that.

PATRICK: I would like to make a hypothetical case if I may. Assume for the moment that I'm a university librarian on a campus and I have my card catalog split up by department: physics, chemistry, biology, and what have you. I have subcatalogs in six locations, none of which is more than a mile from the central library. I have a central store of books, which of course has the central main catalog with it, and I want a teletype in each location and six teletypes sitting side by side in the main library. I want them tied together by leased line. The physics librarian, for example, can type in a request on the physics teletype and ask if a certain book is in the stack at the central library. I have six remote teletypes, six teletypes sitting side by side in the central library, six pairs of terminals here, and 6 miles of wire. Please what would that cost me? Can you get it within $10 a mile per month?

I ask this because I don't think the relationship between the hand sets—the keyboard, if you will (consoles we were calling these yesterday)—and their interface has been made clear yet. The communications people can sell up to the wall mount and the wall mount on the other side (there are units on both ends).

QUIRK: It's less than $10; there is a little problem when you go a mile. You are in-exchange.

PATRICK: Let's say its $5. A teletype is about $60 per month so if I had two teletypes per link that is $120. I add $5 for the line, that's $125 per link. I have six of these links so $750 is about what it would cost. So the central facility with the outlying catalogs with leased 24-hour communication is a practical thing today.

F. B. ROGERS: I have another hypothetical question to pose with lots of conditions. Suppose I want to transmit a lot of information to a lot of different outlets in the United States. This is not in response to individual requests; I am regularly sending this amount of information to these outlets. I subscribe to Wide Area Telephone Service (WATS) in Washington, D.C. I have 100 outlets all over the United States to each of which I wish to transmit 1 million bits per day. I want to do this during the 12 hours of the night, which would give me about 5 minutes for each outlet to translate the million bits. I'm unloading a magnetic tape at my end, and I'm loading a magnetic tape at the other end. Is this foolish, is this out of the question for the conditions I stated? And if it isn't, is it reasonable to suppose that I might be able to do that? I believe my WATS line from Washington would cost me around $2,400 a month. What then is the equipment at my end and at these 100 outlets around the country going to cost per month? It's the same million bits that I'm transmitting to each 1 of these 100 outlets.

QUIRK: The fastest speed we have at the present time over a WATS line (WATS refers to the rate treatment and is a normal dial telephone facility) is the 201A data set which transmits 2,000 bits per second. If my mathematics is correct this would take about 500 seconds to transmit a million bits, or about 8 minutes total at this speed. Of course you would have setup time, answer time, and so on which would probably average about 14 to 16 seconds per call. So it is within the range of possibilities. Now if we use the 202 set you increase your transmission time but you cut your costs about a third, as far as the data set is concerned, and this is probably about the lower limit.

F. B. ROGERS: What does the 201A cost?

QUIRK: The terminal cost is about $70 for the 201A, which is the 2,000 bits per second, and about $25 for the 202.

F. B. ROGERS: So if I had a 201A in each of these 100 locations, I've got $7,000 a month plus $2,400 for the WATS line; so for a total of $10,000 a month I could transmit a million bits routinely day after day to 100 locations in the United States.

QUIRK: This is a million per location, or did you mean a million total?

F. B. ROGERS: A million per location. In other words, for $10,000 a month or about $100 per installation per month I could transmit this quantity of data to each of 100 locations.

EMLING: I'd like to point out that this is tight; 8 minutes each times 100 locations is 800 minutes, which is more than your 10 hours. It will take 2 lines to do this for 100 locations.

F. B. ROGERS: If I have Wide Area Telephone Service, I have it 24 hours.

EMLING: If you do it 24 hours, all right.

DUBESTER: What is the speed of the 202?

QUIRK: The 202 is 1,200 bits a second.

PATRICK: This assumes the tape transports at both ends?

QUIRK: I have not taken into account any of the gear except the telephone terminal costs.

F. B. ROGERS: I don't want to know how much the tape transports at each end cost. One other question. Does your previous answer about the average error rate of 1 in 10^5 bits apply to this situation?

QUIRK: This is correct.

WILLIAMS: I want to be sure that I understand your answer. Are you saying that he cannot transmit from one station to 100 other stations simultaneously, or he has to do this in sequence? You can't transmit from 1 to 100 simultaneously?

QUIRK: Not on a WATS line.

WILLIAMS: Is it possible to do it with some other equipment?

EMLING: Yes, but you wouldn't want it. This would take a very special setup.

EDMUNDSON: First of all I want to say that I'm sure Emling and his colleagues need no defense on my part, but I don't think we should bring our problems and expect definitive solutions in a short time. I do think the important point that we should remember is that this is probably the first time that librarians, automation people, and communications people have started to address themselves to these problems. The solutions are not apparent, the alternatives are many, as has been pointed out time and time again by Emling. Again we come back to the question of total system costs. What are the alternatives? What criteria are we going to use in making the final decision?

I hope that the librarians leave this symposium with the feeling that they haven't been given a lot of slick answers by the computer or communications experts. At the next meeting perhaps we can speak a common language, which we certainly have not had up to this meeting.

CLAPP: There are many factors to consider in

735-898 O—64——16

communications. A study was undertaken at the University of Michigan some years ago to cost out the possibility of abolishing the departmental library catalogs on that campus and installing electronic equipment by which the central library catalog might be consulted. These were the costs given us. (The high costs were mainly due to the queuing problem in the departmental libraries.) It would cost $300,000 to install a system of the kind desired; it would cost about $50,000 a year to maintain the system. The savings, however, from abolishing the departmental catalogs would amount to about $10,000 a year. Now you can't persuade a university administration to go in for a system with costs like these, even though the service to the users of the departmental libraries would be improved. It might have been that if the service to the departmental users had been improved by a great many orders of magnitude, you could have some hope of persuading the university administration, but the improvement was not enough to justify an investment of that cost.

Let me give you another example; at the University of Virginia some while ago they installed a closed circuit television system by which the departmental libraries could consult books in the main university library. This worked. Unfortunately, it was a poor choice of a test, because the departmental libraries were so close, a matter of 100 yards or so, from the university library that no one would not prefer to walk over through the beautiful autumn air to the main library to consult the original rather than to examine an inferior facsimile on a television camera. This persuaded us that distance was a critical factor.

We then costed the thing out in a municipal situation with a 6-mile difference through crowded city streets between a research library and a potential user of the research library. Here, however, it turned out again that the cost of maintaining the transmitting system and the costs of the hard copy at the receiving end was far in excess of a simple Xerox operation at the sending end and a motorcycle or a truck running at regular hours.

So again I raise the question which Edmundson reinforces. We must do some further study of this business on a purely economic basis to find out what we wish to get advantage of and how we get advantage of it. If this discussion merely opens up that area, it seems to me it has done what it needs to do. At this point I will say that if I'm 2,000 miles away from Washington, I'm willing to pay $1.75 to get a set of cards in one day.

R. D. ROGERS: I just want to add that I think the papers that we have had today have been excellent and that we are deeply indebted to the authors of those papers. Mr. Emling, for you and your group who are not trying to sell anything, I would say you did the most magnificent job of soft selling that I have ever seen. Thank you.

SECTION VII

The Automation of Library Systems

The Automation of Library Systems

GILBERT W. KING
Itek Corp.

Defining a System

What is a system? It is a mathematical model of the operations required, together with a detailed exposition of the implementation of the necessary functions. The mathematical model is needed in order for everyone to have a very precise and formal description of exactly what is to be accomplished. Too often such formalization is not forthcoming, and the alternatives of an automatic system are presented on shifting ground, so that the exact objective is left vague. It is true that an automatic system will provide many attractive and unusual features as byproducts, but the system must stand on one set of primary objectives.

The plan of implementation too cannot be casual but must honestly set forth the system as a whole. Too often a system is approached piecemeal. One feature is implemented, and then the next difficulty is attacked, with the result that the so-called system is a sequence of black boxes of excessive cost. A system, in terms of equipment, must be integrated, which means there is a great deal of feedback from every element to preceding elements. A good rule of thumb is that each device should be responsible for three functions. A system, like a chain, is only as strong as its weakest link, although a system can be a network of chains. This weakest link must be identified clearly, and the major technical effort expended in making it as strong as possible and, where possible, in creating alternative linkages.

In an area such as library operation a mathematical model is hard to establish. Mathematical models usually are highly quantitative, hence arithmetized, expressions of the functions required. There is nothing arithmetical in library operations, so the mathematics must be in relatively unknown fields, specifically in set theory. One must be careful not to confound the model with arithmetic just because computers are the popularly proposed tool of automation.

It remains a major task of those working in the field generally called information retrieval, of which librarianship is a part, to construct detailed mathematical models of the aspects of given retrieval objectives. The implementation of any objective cannot honestly be proposed until such detail has been written out and approved by the operators and users. The current confusion in this field of endeavor can be traced back to this fundamental lack of definition of the problem.

A Mathematical Model

A library is generally considered to be comprised of a set of discrete items: books, manuscripts, journal articles, etc., say s_1, s_2, forming a set S. Each item contains a miscellany of statements (sentences) σ_1; σ_2 . . . A patron of the library comes to find an answer to a query, which we can simplify to the following question:

$$Q = \text{Is there a statement } \tau_i\text{?}$$

The statements τ_i may be very specific pieces of information, or they may be oblique statements relative to what the patron really has in mind or to what he will appreciate as satisfactory. We assume that the patron will examine different oblique statements τ_i until he reaches a satisfactory answer or is convinced nothing is forthcoming.

In the simplest queries, τ_i is identical with σ_j, that is, there is such a statement actually in an item in the library. The search problem is merely one of matching, $\tau_i = \sigma_j$. In general, exact matching is not possible, and some more general equivalence must be established. For example, τ_i might be in English and σ_j in German. More frequently τ_i will be in one English construction and σ_j in another. No one as yet has discovered a set of transformations T such that

$$T\tau_i=\sigma_j,$$

i.e. so that the statement (or sentence) τ_i can be transformed to match identically any other form of the idea, in particular the forms which actually might appear in the library.

Current use of a library is based on the principle that there is a transformation, or more precisely, a mapping of τ_i on some primitive Π composed of the words (or phrases), π_k, of the library catalog cards. It is assumed that the librarian has beforehand mapped all the contents σ_j of an item onto the same primitive index language. The awkwardness of libraries lies in the need for the patron to discover the mapping used by the cataloger, and the fact that the mapping of both the patron and cataloger are many-to-one. That is, all the statements in an item have to be mapped into the title, or at most a few subject headings. We shall not discuss the possibility of more elaborate mappings or the ultimate objective of finding the instantaneous transformation T (such that $T\tau_i$ can be identified with a σ_j), although any proposed system must be of such a nature that it can develop, without revolutionary changes, to accommodate any steps in this direction.

Our model then consists of the proposition that there is some mapping M, commonly called cataloging (or extended indexing), and that

$$M\tau_i=\pi_k \text{ or null}$$

$$M\sigma_j=\pi_k \text{ or null}$$

By null, we mean that some statements cannot be cataloged or indexed.

Further, a library implies that the descriptive cataloging of an item is represented by a set of symbols s_n such that

$$\pi_k\subset s_n.$$

This means that descriptive cataloging of item s_n must obviously include the mappings of all the statements in it. For example, the title *The Face of North America* implies "the Goosenecks are in Utah." Such, in fact, is not the case. All statements of geology and geography of North America are not in this particular book (even those one might expect). Furthermore, there are many statements in a book which one would not expect to find there. Because of the loss of information in a many-to-one mapping, catalogers cannot possibly provide answers, τ_i's, even for all anticipated queries, and on the other hand, a great deal of information σ_j not included in the same item, s_n, will not be retrieved. (We are not neglecting the fact that M will not be specifically written out but largely lies in the reference librarian's head.)

This is the situation, for better or for worse. The operation of a library, then, is described by the mapping M. This mapping is an elaborate and heterogeneous structure, but, nevertheless, in a first-class library is very specific. It can be learned and can be well executed—even the patrons can become thoroughly conversant with it. Nevertheless, no amount of human effort can overcome the inevitable loss of information in a many-to-one mapping.

In one sense the mathematics of the mappings we here encounter is very simple. The basic mapping can be expounded as a table, of the form $x : f(x)$. Here x is a string of characters (letters, numbers, punctuation marks) consisting of the references of the items in the library. The function $f(x)$ onto which these strings are mapped are carefully selected by competent catalogers.

The basic table is the author file, a list of all items identified by author. Secondary identifiers are given: title, edition, etc. In the use of such a file or catalog it is assumed that the user implicitly knows that an author writes sentences about certain types of information. The title subclassification divides the author's sentences or statements into groups (although all of us have had the experience of finding a quotation in a different book from the one predicted). In short, an author, in his writings, establishes a subset of σ_j. This in turn implies a subset τ_i to the research worker who knows more or less vaguely a mapping $M\tau_i$ equals $M\sigma_j$ equals author k. This is how we actually use a library most frequently.

Many times, however, the user does not know an author and must look under some title or subject heading. In this exercise, he assumes some m' (the more he uses a particular library the more m' approximates M) such that, hopefully $m'\tau_i$ = some $M\sigma_j$ = some π_k. This equation expresses the exasperation of the patron in trying to find a mapping of his ideas equal to the mapping of the librarian who made up the subject catalog.

In the above mappings m (or m') there is necessarily a great deal of subjective judgment, e.g. is the book about the history of Europe or the history of England? There are, however, some fine details, not of such a sophisticated nature, but complex and necessary. There are relatively trivial, but necessary, "mappings" of pseudonyms into standard author entries; of synonyms into approved subject headings; of cross-references, and the like. Thus, the alternate mapping of a σ_j, or a τ_i into a π_k may go through a sequence of intermediate mappings. The efficiency of a library catalog and the assistance given to the patron determines the adequacy of a library. Automation can, indeed, not only elaborate the types and structure of all these intermediate mappings but also speed them up, so the search trail goes at a satisfying speed to the patron and provides him with clues to assist himself.

To be more precise, each particular or intermediate mapping consists of a table whose entries have five parts.

$$q_i s_k P^{\alpha} s_l q_j$$

Here q_i is the name of the table being used, e.g. authority file; s_k is the set of symbols to be mapped, e.g. the tentative author name; P^{α} is a tape shift control instruction, where α indicates the amount the tape is to be moved; s_l is the mapping, e.g. real name; and q_j is the name of the next table to be examined, e.g. official author catalog. A search through a sequence of table entries of this basic type, then, step by step, converts the initial sequence of symbols s_k, say through intermediate sequences s_k', s_k'' . . ., to the output sequence s which is the best the library can do. Hopefully, it is the call number of an item which does, indeed, contain a statement of σ_j, equivalent in the sense of the statement τ_i in the query.

The model is then a set of tables of substitution rules.

Required Functions

Search.—This mathematical model can describe precisely what goes on in a library search, and, indeed, can describe search procedures of greater sophistication than any now in existence. It therefore describes the functions which need to be implemented. These are the abilities to:

1. Find table q_i
2. Recognize a sequence s_k in the query and match with an entry in table q_i
3. Replace the string s_k by s_l
4. Refer then to table q_j

Thus, as far as functional requirements go, there is only one "algorithm"; it is composed of the above four parts. This simple algorithm is used for every type of mapping, however heterogeneous. The mathematical function executed is substitution of the one string of symbols by another. These substitutions are transformations but of a nature that may, and indeed do, transcend any arithmetic or logical transformations. Substitutes are far more powerful and encompassing than any set of algorithms of a conventional computer.

Generally speaking, the substitutions are arbitrary, in the sense that the substitution of "Mark Twain" for "Samuel Clemens" is not computable by any other algorithm (arithmetic, logical, or heuristic) as would be the case of substituting 2+2 for 4.

The single algorithm of searching in a table forms a deceptively simple mathematical model. The essence of the problem and its complexity lies in the details of the entries themselves, and there is indeed no limit to the sophistication attainable by these tabulated substitution rules. (For further amplification of this point consult works by Alan M. Turing and D. M. Davis.)

A central feature of the library problem is that essentially all the tables are extremely large, relative to numerical data processing. As the search strategy is improved, this number will increase—to some extent by decomposition of the larger tables, e.g. the subject index.

We have already determined the basic physical properties of the system: rapid search by an intrinsic address in memories of extremely high capacities for an entry, with provision for substitution of characters whose numbers are not equal and for insertion of the next table number q_j.

It is seen that technical difficulties encountered in computer systems have been bypassed; for example, impedance matching or interface problems. This is because each successive step in the processing by substitution is identical and can use the same equipment. All that is required is to locate a succession of different tables all of which can exist in the same memory.

We have, therefore, achieved a very desirable feature of any system, namely basic simplicity and uniformity of the equipment. This in turn permits modular construction—as the system grows in size or sophistication, identical units may be added without any change in system operation or programming.

Multiple Use.—So far, we have considered the individual patron. It is a characteristic of automatic systems of this kind that, in order to perform in a nontrivial way or in a way that cannot be emulated by humans at lower cost, they must become large and therefore expensive. Not only is the equipment expensive, but an equal amount must be spent on preparing the contents of the system, and another substantial amount on loading the system. The total cost is prohibitive unless the system is to be used simultaneously by very many users. This feature in turn calls for additional expense, and unless the marginal costs are small the whole project is out of the question.

It is therefore necessary to provide several users simultaneously with access to the whole system, without mutual interference or serious queuing. This imposes a requirement on speed in the basic and central lookup process and necessitates buffering of the data of the search path for each user. It also requires communication and switching components in the system.

Serials.—The most troublesome items to handle in a library are the serials, but the difficulties are amenable to mechanical solution without the need of advances in methodology of the sophisticated nature required in the general search strategies.

Some of these complexities could be resolved by the use of consoles operated by the library staff. One problem arises from the variety of ways serials are titled and by the numerous abbreviations. Clearly, these complexities could be resolved by automatic reference to a file, or table, of all the variants. The output would be a standard form, or indication of ambiguity, e.g. *Am. J. Phys.—Physics* or *Physiology?*

The lookups on abbreviations, giving the standard form, could simultaneously refer to another file in which semipermanent information relating to the serial was kept; e.g. previous titles, number of issues per year, etc.

In the case of serials, there is more changeable information and updating (of accession) than occurs with other items. For these functions easy writing memories are suitable and conventional business machines appropriate. Indeed, the function of keeping the serial collection cataloged is similar to, and overlaps, the business accounting functions of the library, which can be handled in a conventional way.

Many observers of large libraries are concerned by the fact that serials, especially journals, are not indexed in depth. For example, *History Today* is quite an inadequate title to serve a student of history, since this journal covers a very wide variety of topics and, in fact, to the querier could be a source of many topics it never has covered.

Indexing such material, even by author and title, is beyond the present work capacity of a large library. To obtain such coverage the big library system should be designed to incorporate all the indexing done by smaller libraries and professional societies.

Graphics.—Part of the library system will be the storage of information in graphical form, i.e. photographs of the printed page, with or without pictures, presumably in greatly reduced size. This parallels the more conventional storage of items, such as books, journals, newspapers, etc., in their original physical form. Use of graphical storage is growing, and therefore complete integration into the system must be made and in such a way that future expansion can be accommodated. There is no problem in recording the geometrical position of an item in the graphic store in the same fashion as the shelf position of a book.

However, graphical storage, especially in microform, has advantages over physical storage. Rapid access to the items is possible so that immediate display to the user on a console may be made, thus providing him with a new system function—browsing. The console also can provide a new function by supplying him with throw-away hard copy for his personal use.

The system implications put certain specifications on the graphical storage. First, this storage should be able to grow indefinitely without decreasing the rate of access by one or several users. Second, access for browsing must be below the psychological irritation level of one-fifth of a second. Third, the microform must be such that a high-quality display of sufficient brightness can be made at a console.

For the long run, the system should be designed so that the graphic storage can be used to assemble a complete document at the output stations, giving the user a tailormade booklet containing an exposition in answer to his query.

Digital Storage.—Some of the data in an automatic library will be storage in digital, i.e. machine-readable, form. At first, this type of storage will be limited to the card catalogs. As indicated elsewhere this will permit pieces of the search trail to be executed automatically within the system. To make use of human intervention, the digital information will from time to time have to be converted to a graphical display at a console. This may happen at any point in the search trail so that the coding of the digital information must always be such that a full display may be made with adequate fonts, styles, and symbol sets.

In the incorporation of these functions into the system, it must be borne in mind that, in the long run, more and more data within the system will be stored in digital form. For example, all publications printed on machines operated by paper tapes could be stored in digital form, if the tapes were made available to the library. It is true that, at the present time, our methods of automatic search are not developed far enough to make use of such storage of full text in digital form, but there is no doubt that such methods will become practical in a few years. Potential expansion of this form of data storage and processing must be part of the system.

Principles of Design and Choice of Equipment

Stating the requirements of a library and then constructing a mathematical model which evolves into a functional model of the system lays the groundwork for consideration of equipment for implementation. When we are faced with the selection of existing equipments or the design of feasible new equipment to meet the stated requirements, we enter a phase of feedback. That is, the limitations (and the cost) of existing or feasible equipment may be reflected in modifications of the statement of requirements, even though the latter were stated against a background of what is practical.

The principal equipment components will be examined from this system point of view.

Storage Capacity and Accession.—The library problem arises at the point when the volume of material exceeds the capacity of a reasonably sized staff of human beings. There is no point therefore in considering automatic systems which do not have capacities superior to a typical library staff. In other words, the library problem only arises and is only worthwhile solving when we wish to deal with several million items. Otherwise, manual methods remain superior, e.g. experience at Battelle, etc.

This consideration eliminates many types of storage devices, leaving only those with capacities in the hundreds of millions of bits.

Coupled with capacity is the requirement of speed of access:

(a) To provide real-time lookups for an individual search trial.

(b) To provide real-time performance to many users at the same time(for a library cannot serve only one at a time).

(c) To encourage growth of use and thereby to distribute the capital and the loading costs of the system.

The solution of the capacity-access problem requires a system design study concerned with trade-offs of parallel and sequential search means, queuing of lookups, etc.

Interface Problems.—If a system is designed in such a way that it requires a large variety of components, the costs of construction and of maintenance are high. Furthermore, new system costs arise because of interface problems. The system should have as few different types of units as possible, a consideration which should reflect back into the mathematical model and its functional implementation. An attempt was made along these lines to characterize all the wide variety of processing in a search procedure to one algorithm—table lookup with substitution rules.

This has the effect that only one type of memory, with one type of addressing, is necessary for all the kinds of data processing required functionally. In terms of equipment, only one type of memory and electronics is required in the system, the various functions being executed by appropriate types and organization of data in the memory. That is, the heterogeneity is handled by the form of the data rather than by variety in equipment.

A system almost by definition must consist of different units, interconnected. However sophisticated the memories and the data within them, the library function has to include human beings—the users, reference librarians, descriptive catalogers, classification experts, etc. In the interests of system simplicity, an attempt has been made to accommodate all these interactions with one type of equipment—the console. These consoles may possibly vary in complexity according to the operator's functional requirements, but the selection or design of consoles should conform to the principle of standardization.

Data Transfer.—As long as there are several kinds of equipments, especially if these are distributed in a large building, consideration of the transfer of data from one unit to another must be incorporated early in the system design.

For example, it is obvious that all units should operate with the same code structure. Six-bit codes with prefixes (with 1-bit error detection) should be decided upon as being about the simplest in terms of device components, yet capable of providing, with little loss of throughput rates, for the full spectrum of symbols needed in a library.

The data rates in the systems visualized are not high in terms of the communication industry, and every effort should be made not to have any part of the system demand a data rate which deviates largely from the average. In this way, the linking of the various units of the system in a large building can be done with relatively low cost of wiring and minimum complexity in converters on and off the wire.

Communications.—This concept of linking many libraries and other users together for mutual support and cost reduction raises a different set of data-transfer design considerations. Probably high data rates, in bursts, are desirable in order to avoid the cost burden of having long distance communication channels open for long periods of time. Compatibility with the local data links will certainly not be insurmountable but must be planned far ahead even though the actual realization of the network feature of the system may come at a later date.

Terminal or Output Devices.—An obvious, but generally neglected, function of a library system is to provide information as an output. The consoles have been discussed as aids in the search trail, but they must also serve the function of displaying the termination of a search trail—the information the man came for. In many cases a visual display, of the same kind used in the search itself, will be sufficient. There is an additional requirement that the console have some means of supplying the user with an answer to his search in the form of readable material; that is, low cost, throw-away hard copy of selected items which were displayed to him on the console. The important point here is that this requirement must be integrated in the system and not left as a last-minute addendum. For example, it is highly desirable that such output be easily readable. Thus, although the user might tolerate relatively poor quality from the printing point of view, he would want a variety of fonts and symbols. The console and its printer must have a capability of responding to this richness of symbolism, which, as we have already pointed out, must be preserved and exist in the system as a whole.

The above remarks have assumed that the result of a search is a set of items displayed or printed at the console, for example, bibliographic references. Generally speaking, the true termination of a search is a physical item in the library itself—a monograph, serial, etc. The terminal display would actually be the shelf location of the item. There is now the option of having the user at the console set in motion the mechanism to provide him the physical item automatically or of using a separate call system.

In the case where the item is stored in the library in microform (newspapers, for example) clearly the graphical display should also be presentable on the console. This is a third functional requirement of the console, and thus its design again must be looked at from a total systems point of view, even though all these features may not be implemented at the outset.

There are several other outputs available from an automatic library system. An extremely important function of the Library of Congress at the present time is the provision, for the country as a whole, of descriptive and subject cataloging of most new books (as well as a variety of older ones acquired from time to time). In principle, this cataloging can be used by many other libraries, and in order to reduce their cataloging load, it is imperative to disseminate the LC cataloging in the form of printed catalog cards as quickly as possible.

In order to do this, the system must have output equipment to print and disseminate catalog cards. Furthermore, these cards must have the high quality, multiple fonts, and range of symbols available on current LC cards. Data in the form of adequate coding, to meet this requirement must exist throughout the system as a whole.

Another output, with almost identical requirements, is the publication of the National Union Catalog. To this we could add the publication of book catalogs of the holdings of the Library of Congress or of its divisions, or of associated libraries in the network.

Looking to the future, we could expect the automatic system to preserve and to accommodate various bibliographic searches, to assimilate and coordinate these, and to provide published bibliographies as an output.

Large vs. Small Systems.—There can be little argument with the general description of the automatic library system outlined here, and moreover the details can be defended on the basis of the present state of the art. It is obvious that such a system could give the type of performance generally required, especially in being open to inclusion of new methodologies. The large system has an intrinsic value in providing a market in which the manufacturing costs and distributed expense of system design can be kept reasonable. After its installation, the concept of a communications network linking smaller or less automated libraries having only a console or two should become attractive to many libraries. However, this is a concept of the future.

One often hears this question: Why cannot automation proceed now on a small scale utilizing business machines often available in or near libraries? This is a legitimate question, but the main arguments against this approach should be apparent from the preceding discussions. First, there has to be a realistic examination of the capacity of any interim automatic system. Has it memory capacity for all the material of the library? The emphasis is on "all," because a partial system is sure to be inadequate and cause frustrations. Can the data to be processed be inserted in an interim system without formatting and loss of valuable detail? When a better system becomes available, does the input have to be done over again? Can the interim system provide the patron with a rapid response and provide real-time guidance in his search sequence?

From the theoretical or methodological side we see there are two basic requirements of the system:

1. Capability of storing vast quantities of information in organized groups—tables.
2. Capability of continuously organizing the stored information and incoming information in new tables to increase the power of search strategies.

However good and enticing the theoretical model and its potential methods are, they remain academic until a practical means of implementing them is available. We simply cannot overlook or postpone the solution of the problem of getting information, not only in its raw form, but also in its organized form, into the system.

Too often the practicality and cost of loading data into the system is ignored, and as a result, the sophistication of many proposed schemes must be abandoned in favor of a simpler solution. The loading problem for library systems is characterized by the fact that the data are very heterogeneous. In descriptive cataloging, each item has to be subjected to a fair amount of intellectual analysis and many demand great professional skill. For large collections, indexing in depth by manual methods is out of the question as a preprocessing or input feature. Formatting of the input data, to simplify input or retrieval, should not be accepted by libraries, because information is always lost when put in a format straitjacket. These points are discussed elsewhere in this conference.

The system makes sense only when it includes the type of input described here, where a capability is demanded for preserving all the information of the input data, at a practical volume level and at reasonable cost. This has to be done at the sacrifice of preprocessing and analysis. Ultimately, in fact, the value of automation will be precisely in allowing us to overcome much of the information loss by which analysis is characterized.

For retrieval, however, it is mandatory to have some analysis, even if only at the minimum levels of author and subject indexing. Even these levels of analysis are impossible without some internal processing within the system, for example, reference to the author file.

Some of these analyses, such as indexing by author, have to be—and can be—done for all items. Others, even subject analysis, cannot be done thoroughly for all items, but the system must provide some guidance for the patron; for example, it must index each item under at least one subject heading. It should also provide for a growth of analysis, such as extensive crossfiling. It would be too costly to do all the desired crossfiling at the input stage. Furthermore, subject analysis grows and changes with time as the public's interest waxes and wanes. The system must provide for a dynamic analysis, even at the level of subject classification.

A characteristic of any large store of information is that it is impossible for a reasonably sized group of professional librarians to analyze it in any depth at all. On the other hand, if the library serves any need, it will have users, and no one could be better analysts of a collection than those who use it. The system should therefore be designed to exploit and preserve searches made by the various users. For example, every bibliography made should be preserved in some form by the system so that future searches can be expedited. The system should behave as far as it is feasible like a reference librarian.

No one today pretends that a system can be devised that could answer every patron's query purely automatically. It is for this reason that displays at consoles are an essential ingredient of any automated library system. Although the prime purpose of the console is to assist in structuring the search route, by allowing the user to interact and to guide this route at various points as displayed, the dynamic nature of types of queries asked over the years make it desirable that the various search trails be recorded and brought forth as suggestions when appropriate.

Again, we can remark that the display at consoles is solvable, as far as technology is concerned, but from the system point of view the characteristics of the consoles must be defined to permit integration of these system requirements, over and above merely serving the patron on the basis of a single use.

Programming

To provide all these functions, not only is equipment needed, but it has to be exercised. The data contained within the system must be made to flow, and this is accomplished by what is generally called programming.

In discussion of the console search, we indicated that various files are examined and displayed to the user who more or less directs the search route. Behind these phenomena must be a program that causes the system to implement the demands upon it. There are many "users" of the system—the patron, descriptive cataloger, classification expert, publisher of cards and catalogs—all requiring programs to carry out their needs and desires.

A fair amount of effort has been expended on the programming of numerical data processors for handling lexical material. Quite obviously, the match between material and method has not been good, and consequently a large body of programming for such data has never developed. The lack of response of textual data to numerical and logical algorithms is now widely recognized and a new approach called for.

For our mathematical model, the program can be based on table lookup methods, since these are in accord with the nature of lexical material. Basically, table lookup is the original method of processing data. Only recently has the deep mathematical nature of tables been recognized, especially as a means of making the fundamental approach of Turing *et al* practical. This is what is attempted in characterizing the library problem by a mathematical model based on set theory and semigroups of substitution rules.

Apart from being fundamental, this approach to data processing is very much simpler, perhaps by a factor of 200, when it comes to the details of programming.

Whatever methods of programming are used, certain comments are of interest from a systems point of view. First, the task is quite formidable and justifies all the individual experimentation now going on. Nevertheless, as a community we must recognize the enormity of the ultimate task and do our best to build up a pyramid of programs in the sense of one level of programming leading to greater sophistication in the next. It is too dangerous to insist upon or freeze on a "programming language" for lexical material, but some sort of clearinghouse will soon be essential in view of the manpower requirements.

The approach to programming of lexical material should be divorced from the history and experience with numbers and Boolean algebra. As an example, there is a fundamental difference in writing a program to solve a problem numerically and writing one to handle lexical information. In the former, the whole sequence of instructions must be written out, anticipating all contingencies, so that the machine can come to its answer without human intervention. This is not the case for library problems, because we know such anticipatory programming is beyond our understanding of the problem. Human intervention at all stages of the search trail is necessary and by no means undesirable for, in the process, the patron is educated. This means there is a great deal of independence of subroutines.

Exhaustive anticipation of all contingencies, then, is not nearly as necessary in programming for libraries. Neither is the occurrence of errors at all catastrophic, whether they be program errors, typographical, semantic, or the like.

These attributes of lexical processing have had a great influence in selecting table lookup methods. Tables are made of entries which are mutually independent except for the tracer symbols q_i. Automatic diagnostics are extremely easy and efficient in processing by tables. Heterogeneous operations can be executed by the same table lookup algorithm.

There remains, however, the fundamental and inescapable task of making up the tables. From what has been said it should be apparent that the table lookup system lends itself to evolution. One does not have to code the whole program in all its sophistication to get going. Simple tables, e.g. author files, are basic, useful, and already available. Users themselves, by their searches, can contribute new ideas and methods which can be introduced without the result of a patchwork.

Nevertheless, the sheer volume of data needed to make the basic tables presents a loading task which is formidable in detail as well as in cost. The basic principle is to get material into the system without loss of information, in anticipation of processing which will organize it more and more as the system develops by usage.

Costs

The library problem can be summarized by saying that a technically feasible solution providing improved service is at hand, but the costs are high. For a large system, the cost of automation is about the same as for a manual system meeting the growth predicted in the near future. One could expect, however, an enormous hidden payoff by an increase in the number of users and the provision of more pertinent and timely information.

Nevertheless, it would be highly desirable to get the costs to as low a level as possible. Primarily this can be done by creating a reasonably large market for the equipments. This can be done by proposing systems whose components are general purpose; e.g. quite flexible and adaptable to the different kinds of library services desirable at different places. The equipments should also be designed to adapt to similar types of information retrieval or intelligence systems not normally considered as libraries.

In all the basic equipment areas—of memory, consoles, input conversion, and communications—it seems that the specifications for library systems can be formulated to meet the varying needs both of different libraries and of related information systems. Nevertheless, the system designers, when it comes to details, must keep these needs in mind.

Another way to reduce the effective costs is to increase the traffic through the system. This will be done by the increased services and performance, in particular, by the rapid response. It can also be done by broadening the class of users, particularly by linking many existing large libraries into a network through communications channels. This does not mean that every library in the network has to be equally automated. Rather the automatic services of the initial or central library should be made available to users at the terminals of the network. Thus, the more large libraries become automated and mutually connected by the communication network, the more service the total system will provide. We may then indeed expect a tremendous growth in library services as part of our national culture.

Conclusions

Although there is very rapid development of both equipment and methodology desired in automatic libraries, it is not too soon to begin work on the system design. A good system can be specified by repeated cycling through these levels: definition of functions; development of a mathematical model, and a functional model; implementation and programming by equipment. Throughout this process system designers should keep in mind the areas which must be open-ended to accommodate future technical advances and customer adaptation.

It is recommended that the design of a system for a large library be begun now, based on the studies outlined in this conference.

CONFERENCE SESSION VII

An Experiment in Communication: Introductory Remarks

BURTON W. ADKINSON
National Science Foundation

The National Science Foundation was delighted to be able to participate as cosponsor of this conference. I consider it an experiment in communication. It was our desire to try to get both leading librarians and technologists together to discuss the present status of library automation and indicate the paths we should follow in the future. I have a selfish interest in this, because in the National Science Foundation, and in particular in my office, we have to find answers to such questions as the following: What efforts should we try to assist? Where should we put our emphasis and support? In what areas should we work? This is always a problem. There are a multitude of places where money can be spent.

There is another experiment going on here. I listened to it last night and smiled to myself, noticing that the title of the discussion was *Communications*, yet I noticed that on many occasions the communicating wasn't going so well. We didn't quite understand one another. Now this is not even unusual among a group of librarians who work together all the time. When you get technical people and librarians, each group with its own jargon, with words having a peculiar meaning to each, you can have trouble trying to understand each other. We need more discussions of this type; I think it will be helpful on both sides.

The National Science Foundation is interested in assisting in this field. Our emphasis has not been on libraries because we felt that the tools that libraries are getting from the fields of science and technology, i.e. monographs, indexes, abstracts, and some of the other printed tools, have been far from adequate. Our great effort in the past several years has been trying to upgrade these so that they will be first-class tools. They were definitely in second and third class several years ago; most of them are much better today; many of them still have a long way to go. The compilers of these tools are worried about the same problems that you are worried about. They are also asking themselves what kinds of service they should give. How should they package their materials? How should they use these new electronic machines to further their work? They are experimenting in many different ways, and I think that they have learned to use the tools, some of them, in handling routine repetitive operations.

As far as relieving anyone of intellectual effort, I don't think that, to date, the machines have; on the contrary, they have increased the need for greater intellectual effort. It is my prediction that this is going to continue, and the introduction of machines into libraries is going to demand higher intellectual effort on the part of the librarians because they will be relieved of many of the routine repetitive activities which they necessarily have to do today. I say this with considerable confidence because when I look at the computer field in relation to mathematics, I remember that 10 or 12 years ago, when computers just started, people said that it wouldn't be long until we wouldn't need so many mathematicians. Today we need more mathematicians, but we don't need so many people who can compute. We use the computers for the computing, but the demand for mathematicians is greater today than it was 10 years ago, and the caliber of mathematicians asked for is much higher than before. I think the same thing will be true in the library field, but we have a long way to go and a lot to learn.

Now my job is to get out of the road and let the people who know something about the topic today perform. Our topic is "The Automation of Library Systems"; this is the goal we are looking forward to. Now you can interpret the word "automation" in many ways. Many people think that automation of library systems implies that everything will be automated, but we have to think very strongly in terms of man-machine relationships and make sure that the machine is used where it is most productive and that man is used where he will be most useful. Automation of systems is the problem of how to use both men and machines most effectively.

I am glad to introduce the discussion leader for this session, Foster Mohrhardt.

A Challenge to Habit: Some Views on Library Systems Analysis

FOSTER MOHRHARDT

National Agricultural Library

It is inevitable that in the final session there is an attempt at a summing up. I would like, however, to begin with a quotation from Robert Fairthorne. He has said that we get off the track in this area when we concentrate on "what the machines could do, rather than what they should do. Neglect of the second consideration sometimes allows absurdity to undermine ingenuity." He also said that "Automatic retrieval entails not so much mechanization of the library as of its staff and users, in that it must both manipulate and talk about the documents for them."

Dr. Boutry of the International Council of Scientific Unions pointed out several years ago that, at the present stage of development of documentation applied to libraries and information centers, our major problem is sociological. We have the techniques, we have the needs, we have the problems, but there seems to be an emotionalism that creeps in—problems in habit, difficulties with people—that are really the impedimenta that keep us from moving ahead as rapidly as we should. I thought of this last night when some librarian asked Verner Clapp, "Whose side are you on?" It isn't a matter of whose side you're on. I think we're all here to do one thing and that is to give people better service. There has been evident some division between librarians and the technical people. Speaking as a librarian, I'd like to defend the position of the computer people. They have all come here with the purpose of being helpful; this is a cooperative venture.

Areas for Discussion

There have been some major elements, at least ones that I consider major, that have either not been mentioned or have been touched on only briefly during this conference. I think we should consider them today.

We should pay much more attention to costs than we have. Copyright was mentioned only in passing. We ought to clearly determine the time scale we're talking about; are we talking about activities now, or 5 years from now? I think we ought to consider whether we're discussing equipment for individual libraries or for groups of libraries. We ought also to recognize that when you move into these systems, you probably won't make them retrospective but will start where you are and move on into the future. I think also that in trying to determine the needs for automation, we are not concerned only with size and cost, but we also have another major interest: our users. If we have users who are making complex demands on us, this might be an equal consideration with the size of the library and the complexity of the operation.

Feasibility Studies

Now in order to bring this more clearly into the realm of the practical, I thought I would outline some of the steps that you might want to take, as administrators, after this conference. As King points out in his paper, there is a sequence. You have to state your requirements. (He recommends a mathematical model, which he'll discuss in more detail with us this morning.) Then you begin thinking about equipment. Now one of the ways that we can evaluate and ruminate and decide what we're going to do is to look at some of the methods that have been used by libraries in studying this problem of feasibility. I'd like to stress that only you, the administrators, can make this determination. You don't let somebody decide for you whether you're going to reclassify your library or what system you're going to use. Similarly, only you ultimately can make the decision as to whether you're going to have any automation.

To aid in making your decision you may call in a consultant or group of consultants to work closely with you and your staff in analyzing needs and recommending solutions. Or you and your staff, calling on those that are near you, can study areas of interest and then turn this information over to a consultant organization for study and recommendations. Or you may consider making a complete self-survey with you and your staff conducting the study.

There are of course many other approaches, but I'd like to give you examples of these three. The Library of Congress, with a grant from the Council on Library Resources, secured a group of technical experts, with librarians as consultants, to make a study of its operation. In the second approach used by the National Library of Medicine, the Librarian, Brad Rogers, selected a segment of the Library that he felt could be improved through automation. He studied it, determined the broad outline of needs, then issued invitations to bid, selected a contractor (General Electric Co.), who then studied the program, designed a system, and is now implementing it. One of the requirements that Brad Rogers laid down—I think it's a very basic one—is that the system must be as good or better than the present system.

The National Agricultural Library Automation Study

I'd like to give you a little more detail about the approach that we're using at the National Agricultural Library, since I'm more familiar with it. In 1962, we requested the Secretary of Agriculture to appoint a Department-wide task force to examine in depth the areas of the library that it felt could be automated and to submit detailed plans for conversion, including procedures, types and costs of equipment, projected calendars of action, staff requirements, and estimated savings. Representatives were appointed from all of the major units of the Department of Agriculture, and we have experts from various scientific areas: entomologists, soil scientists, and so forth. We have a writer, lawyer, accountant, statistician, computer center director, systems analyst, and librarians. In addition we had the help of several land-grant librarians who came in and worked with us.

After a series of meetings during which we tried to indoctrinate the group into the major problems of the field, it was determined that we needed three studies to consider the following questions: 1. What do the research people in the Department of Agriculture want and what system can produce it? 2. What computing system can efficiently handle the library's information? 3. What are the costs of library research under various systems? In other words, this was a user-oriented survey. In order to carry it out, we divided into four working units: one covered system requirements, another systems design, a third costs, and the fourth the writing of the report. The system requirements group was charged with identifying the library users and determining

735-898 O—64——17

their needs; determining the volumes of input, output, and conversion; and making recommendations to the overall task force. The systems design group concerned itself with exploring compatability requirements; visiting other installations; stating, laying out, and identifying computer runs; laying out master tapes; and determining computer time, computer schematics, and personnel requirements. The cost group was to determine the present costs of library functions and those of the information systems in the various Department agencies. (We have a comparable system to that which exists in some universities, for in addition to the main library, we have other smaller libraries or service units operated by the agencies within the Department. We felt that if the Department is to get full value out of this study, it ought to know about the efficiency of those systems as well as its own.) We wanted to know the costs of the proposed systems and to compare these with present costs, and, finally, we wanted to know the total expenditures for information services in the Department.

We're fortunate in that we will have a general purpose computer with a high capability available to us. The work was to have been completed within 6 months, but unfortunately the chairman was called off to do some troubleshooting on another Department project. The report will be detailed and will be made generally available to anyone who is interested.

Now, one of the first things that we found our library staff needed to do in order to cooperate with the people in the systems design study was to flowchart our operations. Those of you who have seen the report on Ed Heiliger's study at Illinois (item 67, p. 139) are somewhat familiar with these detailed flow charts. Ours differ in that we are using a decision-type flow chart in which we describe not only what actions are taken but why those actions are taken. The library staff and members of the task force were given intensive training in logic flowcharting.

While some members of the group worked with our library staff on flowcharting, others concentrated on identifying the users of the library. A questionnaire, based on the advice from experts throughout the country, was prepared by a scientist. We made IBM cards for about 4,400 scientists in the Department; the selection was made by machine of those that were to be queried. We tried to study their fields of interest, their sources of information, and their specific reactions to the library as a source of information. We are trying to assess the role that the National Agricultural Library now plays and also the role that it should play in getting information to the scientist.

I will mention two of the specific studies which may be of some interest. One is a cost study and one is a serial-transit study, which, I think, will compare with the work that Mel Voigt has done out in California. We traced the entire movement of 24,000 serial pieces from the time they were received in the library until they were available to readers. We've tried to determine the patterns of movements, the lag time, the peakloads, and the total processing time. These are findings that are going to be extremely important to us whether we mechanize or not.

A question has been raised about what is needed in terms of manpower to perform a study of the kind we are making at the National Agricultural Library. We estimated that we would need 84 man-months to complete our study, and although we've only used 50 so far, I'm very certain that we will use the full 84 before it's completed.

Cost Studies and Value Judgments

An element in which we are particularly interested is cost. Here I'd like to follow the precedent of some of the earlier speakers and quote from myself. This has appeared before and is as good as I can do now: "Relatively few of us will be able to justify elaborate equipment until we are better informed about the costs of conventional library search and the actual savings which they provide in the total research project. A factory manager can easily justify new equipment that will cut down the cost of $150,000 steel forging. If we are to justify automation in information and library work, it will be necessary for us to accumulate objective data indicating the economic importance of using recorded information in current research studies."

I would like to address myself to this latter point for just a moment because all of our judgments on the value of what we are doing today are value judgments which we make as individuals. There has been a feeling on the part of many that we, as librarians, underestimate rather than overestimate the value of what we are doing. It's about time for us to take a rather strong position, whether we are in government, university, or public libraries, and insist that there be a recognition; that, if necessary, a dollar value be put on the kind of work we're performing. There's an acceptance throughout the country now of the value of intellectual effort. We are derelict in insisting, within our own environment, that library service is as valuable as, or maybe more valuable than, anything that is being done in the university, or as valuable as anything that is being offered by a city or municipality. What we're doing is extremely important and must have at least double its present support. There are two parts to this cost study: first we must know what it costs us to perform the operations that we are now performing, and beyond that we ought to try to make some kind of a value judgment as to the total job that we're doing and its value to our environment and to society.

We librarians all have pressures from people even more misguided than we—administrators, scientists, and others who have heard about this wonderful science information retrieval problem. They are breathing down our necks the next day and they want us to automate so that they can get for practically nothing the kind of service they ought to have. You can't just stand there—I know I can't—and say, "Well I'm thinking about it, but I can do it better the way I'm doing it now." Unless we can prove through studies like the one I'm making now that we can do it better, then we can't tell them this. The very least each of you can do is to know a lot more about what you are doing now and how much it's costing you. Be absolutely certain in your own mind that you can do it better this way, before you tell your administrator, the scientist, or the professor that you already have a better system. Your best approach to this is positive rather than negative.

The Groundwork for the Future

One point I think must be underlined at today's session: we have to consider whether we're talking about planning for today or planning for 5 years from now. Part of the present confusion in talking about hardware results from our not being quite sure just what time period we are talking about. The majority of the librarians here represent educational institutions, and if there is any group that projects and plans in the future, it's this group of university librarians. When you collect manuscripts, when you collect rare books, when you conduct your entire selection process, you are planning as much for the future as you are for today. If you take this same approach in thinking about problems of automation, you will not only benefit yourself, but you will also lay the groundwork for other librarians who will be succeeding you in 5 or 10 years If you don't, they will wonder where you were when this discussion took place.

I'd further like to recommend to university librarians that you take the challenge of these technical experts and give them some of these problems. One, which has persisted for years, is the responsibility you have for supplying reserve reading materials. I don't think this calls on any of the major competences that you claim as professional people; it is routine, time consuming, and expensive. There isn't much satisfaction in it. This is the kind of thing that we ought to ask these experts to solve for us. I think they might do it.

A lot of us have thought about the application of automation to our individual libraries. It must be recognized that even though quantity is a decisive factor in making a final determination, it may not be the quantity that you have in your individual library. I would hope that out of this we will begin thinking more in terms of groups of libraries. Whether by region or whether by type, it really doesn't make any difference because communication technology now enables us to consider them from any standpoint. Librarians, including the group in this room, have pioneered in this. When you set up the Midwest Interlibrary Center, you were thinking in an advanced way for that time. But you must now continue this kind of tradition and think in an even more advanced way about the possibilities of using this new equipment in a cooperative manner. This is an area where the Library of Congress can exhibit dynamic leadership, perhaps by serving as a clearinghouse and focal point to enable those who are interested in working cooperatively to get together to find some solutions.

I think that we librarians should remember this quotation from Vannevar Bush: "We can benefit from machines only if we change our linguistic and clerical habits." I'm as much interested in the last word as I am in the others. One of our major problems is the obstacle of habit, and if there's any one benefit that I had hoped that we'd get out of this meeting, it is a challenge to habit, a little stretching of our minds, an approach that's more visionary possibly than some of us want to take, but it's one that we're going to have to take if we're going to live in the 20th century.

I would like to close by noting that this is the first time that we have had in the Executive Office of the President, someone who has a responsibility in this overall field in which we're interested. I would like to introduce Dr. J. Hilary Kelley of the Office of Science and Technology.

KELLEY: Thank you very much. I want to extend to you greetings from Dr. Jerome B. Wiesner, Special Assistant to the President and Director of the Office of Science and Technology. There are two different groups here, although I think I would not have surmised that by speaking to various people individually. Perhaps you're overemphasizing this difference. I just can't express how wonderful it is for this conference to be, because by having this dichotomy of thought and interest, somewhat like the salesman and the buyer, you bring each other into better focus on problems. I must say I'm very happy to be here.

Mathematical Models and System Design

GILBERT W. KING

Itek Corp.

The Way of the Dinosaur?

When I was asked by the Librarian of Congress to visit Washington and set up the study group, I didn't realize that paleographers were extinct, but I did know dinosaurs were extinct. This was my reaction, a growing reaction over several years, to the big libraries and not-so-big libraries, too. Dinosaurs became extinct, not of their size so much—whales are pretty competitive in size with dinosaurs—but because there was another group of living things called mammals that had a much better way of living and adapting themselves to changing conditions. They had warm blood and they had giant brains, relatively speaking.

Now that was 2 years ago when I thought about dinosaurs, and I'm still of the same opinion. The question is: Why did I and the members of my group stay with it? It is partly because we had a dual role; fortunately, a study of this type is a little different from studying space problems. We have never been to the moon, and it is difficult to know how to do it. But we had all been to libraries and had reactions of various kinds; we'd like to go to libraries some more, and I think our basic feeling was, and still is, that the libraries of this country, and in particular the Library of Congress, are a tremendous natural resource which is not being exploited as much as it could be by orders of magnitude.

One of the observations made at this conference is that librarians, as a whole, don't have this confidence and belief in libraries. Generally speaking, I'd say you lack confidence in what you're doing, in what you're trying to do. This shows up quite a bit in your being so hypercritical about every mention of new equipment or change of habits. This in turn results in the fact that you don't have any research money; no one puts any risk capital in this. Now I'm with a private corporation and we are willing to put in risk capital, but we have to have a market. There isn't one at the moment. We just can't do anything for you, because I don't think you want it. Maybe we don't know what to do.

Now how can we change this situation? One thing that we can do is to look at this problem from the total systems point of view. As has been demonstrated in this conference so far, there is a tremendous amount of attention to the bits and pieces, but no one has seriously talked about putting them all together and having a system. As I think I said in the paper, though I hate to have to say a cliché twice, 2 and 2 does make more than 4, and that's the whole principle of the system.

A Mathematical Model

Now these are all fine philosophical words, but it is not easy to talk about a system. The first thing one has to do is to have a mathematical model, although you may not recognize it as such. I use the word mathematical in the sense that it was used by Mortimer Taube; it just means you have to be very formal in your statements, but this can lead to a misunderstanding immediately. I mean a formalization of the functions you want the system to do, not of the details. For instance, we couldn't care less about the character set at this point. By formalization I don't mean putting in a straitjacket the kind of language you use, or the terms you're going to use, or the nature of the descriptive catalog. I'm talking about the operations of the library.

In a mathematical model one has to be very precise about the nature of the things that are asked for, the nature of the things that are going to be rather significantly responded to. In my paper I used some Greek letters because it just seems natural to us to use Greek letters in a systems pattern. What we try to do is define what we'd like to have our systems do; this is oversimplified but it's a starting point. The material in your libraries is a collection, which I'll just call sentences, represented by sigma (σ). People come into libraries with another set of sentences, and just to make life simple, I'm making them as essentially formal queries of the type: Is there a sentence so-and-so? I represented these queries by tau (τ). So there's a certain amount of homogeneity between these. However, we know that these languages are quite different; that is, I only use about 7,000 words in my whole life, and there are half a million words in the library. So although they're similar in one sense, they're certainly not identical, and there is no relationship between these two things at the present time.

Now what libraries have done is to form another language made up of words, not even sentences, called pi k (π_k), which are their files, subject heading lists, et cetera. This is just to make these statements (σ_j) available to people with these queries (τ_i). The way it's done is with a set of formal rules. You take a book composed of sentences (σ_j) and say. "All the ideas and all the sentences in this book are going to be represented by a few words of this particular language (π_k)." The thing that I want to point out is that this language used by librarians is not the language used in the books themselves; in fact, although this (π_k) might be English, this (σ_j) might be in another language. So this is by no means a trivial correspondence; it has been a great intellectual effort. It is rather fantastic that some 14 million books are under control by some mapping scheme (M) of this kind through all the efforts of the librarians associated.

Now when I go to the library (and I think this is a point that librarians don't understand very well), I use my peculiar, completely individualistic set of 7,000 words, and somehow I've got to map my queries (τ_i) onto this same language (π_k). Although I'm still using English, so the mapping is pretty much the same, what it amounts to is that, over a period of time, you learn to use the language. That's what we mean by learning to use the library. You go to another library; you've got to learn to use that; there's a different scheme of things. That's how we find materials.

Now the information retrieval people are shooting for something that is bigger; in the long run what they're trying to do is to say: If I come in with a query, is there a transformation (T) that will change the query into phraseology so that it could be matched on the fly by scanning individual sentences in the text itself?

$$T\tau_i = \sigma_j$$

This would be output, and there are intermediate schemes of this type. But this is still an unsolved problem.

Now, these are problems that we are going to deal with in any system in the near future, and here again, instead of saying words and writing down some fancy symbols, you have to get down to the meat of what this symbol means. A certain amount of understanding of this has come in language translation. We have come across this type of problem already. What I have in my paper is something that, I think, is a pretty sound way of trying to look at these problems: This formalizing into a mathematical structure can be done with the so-called theory of groups and substitution rules.

The theory of groups is a very sound and essentially new kind of mathematics which is going to be useful for making models in language problems. There are two really good reasons to have a good mathematical model. We have our feet on solid ground when all this kind of automation stuff gets higher review. If we do it right we can have a good solid scheme; it isn't opinionated. The other reason is that this is very exciting; it is exciting to discover that finally out of this chaos, from the scientific point of view, we are getting to see some daylight. I hope that by introducing ideas like this, recognizing things of this sort, we can get some first-class mathematicians interested in our mutual problems. In the last 5 years we have interested one or two, and I'm looking forward to more.

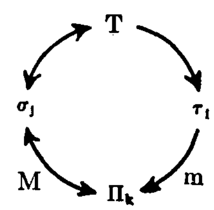
Substitution Rules

Now let me try to explain what I mean by a substitution rule. There are many different examples and we don't have the time to go into this very deeply, but let me try to give one example of a typical situation. Somebody from this audience might go to the library with the following query (τ_i): In the library is there a statement that mechanization of libraries is good or bad? Now he knows very well there isn't any book with that sentence, but, on the other hand, he's got a sporting chance that there is a book that has something equivalent (σ_j). What he finally ends up with is a document called *Automation and the Library of Congress.* Now what's happened is that he's made a transformation of this sequence of letters (τ_i) into this sequence of letter (σ_j) by just plain substitution. He's done it by a dialogue, as it's been called, between him and whatever kind of console or files are in the system. For example, mechanization is a very ambiguous word; it usually means levers and so on. So somehow or another, by a dictionary or, in our system, by the thesaurus of tables of synonyms, it's going to be recommended to him that he change the word "Mechanization" to "Automation"; that is a substitution. If he looked for "Libraries" he is apt to be told, at least in our system, that there are 100,000 books on libraries and that he has to be more specific. The system may give him some suggestions or it may not, but he'll have to scratch his head and say, "I've got to be more specific, how can I be?" And he'll say to himself, "Well, if there's going to be any automation, it's going to be in the big libraries, so let me try the Library of Congress."

This is the kind of thing I'm sure you realize the user goes through with or without the help of a reference librarian. So he's made the substitution for "Libraries," and this is another kind of transformation which can be more or less automated. Now I think it's been quite clear that none of us feel that this is a completely automatic system. We feel very strongly that we have a system where the human being is in the chain of events, because we'll never be smart enough to make it automatic and secondly, because human beings are a lot cheaper than computers.

So this is really all I meant by substitution rules, and the mathematical aspect of this matter is that you can't be vague about these substitution rules either. You have to have tables of these, tables of transforms; these are things that can be worked upon; this is basically what we call programming, once we get the system going.

The following diagram summarizes the previous discussion:



σ_j=statements in books, etc.

τ_i=formal query
Π_k=library catalog, files, etc.
T=ultimately information retrieved by sets of transforms
m=mapping done by user
M=mapping done by librarian

Now another reason for trying to have a mathematical model of this relatively sophisticated type is not only for immediate purposes, but as everyone here realizes, there's a lot of work going on in information retrieval. Certainly progress is going to be made in this area, and we feel that it would be very foolish to devise any kind of automatic system which would not be flexible enough to include information retrieval as a future possibility.

The Library of Congress Feasibility Study

I think no one would argue seriously against the view that the principal point of entry of automation in libraries in the United States should be through the Library of Congress. Now, everyone wants to invent his own system, just as everyone wants to invent his own missile booster, but we are going to have to concentrate on one kind. On the other hand, we don't want to dampen the ideas and endeavors and excitement of the various people who are interested in this. So one of the things that we've definitely had in mind is to have a system that is quite adaptable, because nobody knows the answer now and nobody will even 2 years from now. In this case, we're going to have to live and learn, as the missile people are going to do. Now, the system should be adaptable for other reasons; other large libraries and even the relatively small ones should, we hope, go in this direction and have a similar system. It would save them a tremendous amount of money if the engineering design were done but done in such a way that individual libraries can choose how much of the system they want and can adapt it to their own special needs.

I should make a point that our study considered more than the automation of the Library of Congress; we considered the automation of the library community, in the sense that there is a very powerful group of research libraries. It seemed to us that it would be a tremendous benefit to include all these other libraries in a system so that a great many mutual benefits could be obtained and the cost could be reduced. Our system, when looked upon from the various kinds of applications for users, should be designed so that it is compatible at least as far as hardware goes. If different people have different systems, they ought to be compatible in the sense that they can talk on the same communication network.

I want to emphasize very strongly that our study group believes that, although you might use standardized hardware just to get the cost down, there's no need to go to standardized software so that, for example, everybody has to have a certain classification system. Good or bad as the LC system is, there is no reason why, in an automated system, everybody should use it. In particular, there are specialized libraries, scientific libraries, report libraries, where material is classified in a different way and in different depths. By these principles of substitution rules, we can see how different library systems, as far as organization goes, could certainly live together with each other even if they are connected by telephone lines.

The Transition Phase

Another reason to have the system quite flexible is that we have to face up to the transition problem. Even if we had the money and knew how to build it, we couldn't think of saying that the Library of Congress would be automated by a certain date. This is a stupendous task, and it probably will never be completely finished. We have to work out a scheme so that we can get going without having to go whole hog. A typical information retrieval and library problem is that you really can't do anything better than human beings unless you have a pretty large store. So this transition is not going to be card by card; it has to be in big batches of cards to get started. This is where we have to be more specific about the nature of this class of sentences (σ_j) and this class of sentences (τ_i) in these transformations, and instead of talking about the whole universe reduce this to something big enough to be of interest and small enough to be afforded.

Memory Access.—Now just the way an outlying library might join this system, when it gets going, by getting one console and tying in with the memories, I think too, in this transition period, the Library of Congress could probably start off with one console and one memory. Nevertheless, a principle in systems design is to try to make any major component, such as a memory or a console, multipurpose so that it isn't designed for a very specific part of the system or a specific phase but, at least with only minor modifications, can serve many purposes. For example, the console is multipurpose in the following sense: at first, it would be available for use by the descriptive catalogers or their junior staff; then as we get going, and get more in the memory, reference librarians would use these consoles to help the patrons, and in the next phase, outlying libraries would have a console to communicate to the system. From the point of view of cost and also maintenance, these consoles should be the same kind even though they are performing relatively different functions. The same thing is true of the memories. Obviously we are going to need enormous masses of memories in the long run, and these memories are going to be used to contain different kinds of files. We could have a tremendously complex system on our hands unless we try to design these memories or the access to them in such a way that they will serve all the heterogeneous types of queries that will be made in the final system. And, here again, I think that this mathematical model of addressing memories from the point of view of the substitution rule is a good principle that will simplify and standardize the hardware of the system.

File Conversion.—I would like to discuss how the transition would take place and what the system might contain in the first few years. I'd suggest that the first thing to convert is the authority file so that descriptive catalogers could get to that material more quickly and make the results of cataloging available to other libraries in a much shorter time. Next should probably be the serial records, because these to some extent are not related to the main file in an integrated way such as other kinds of documents are. There are so many unsophisticated problems with serials that are essentially inventory or business problems, and these are things, as Patrick pointed out, that we know how to do and can do now. The next thing I'd like to see converted is the official catalog; *now* I'm really talking about an expensive item, but until that, or a good portion of it, is put into memory so that automatic access and tracings, in my sense of the word, can be made through that file plus the authority file, you won't have much of a system. From the cost studies we've made, it looks as if it would be quite reasonable to start on this, at least by taking part of the file; namely, science and technology. I would suggest starting here because we might be able to get some funds for this.

Of course, we'd like to convert the National Union Catalog, and this is where other libraries would become interested, for this means that they could have information about their own catalogs. Then we have subject files; these at first, of course, would be in the Library of Congress system, but since these large memories are relatively cheap, we could put in different files that other people have worked on. There's no reason why the system shouldn't include the Decimal Classification system, so that people could switch from one system to the other. This avoids the cost of trying to program the LC system into the Decimal system and avoids a lot of irritation. When this is available to some degree, we can start talking about consoles for reference librarians, so that users can start using these files. At first it would probably be sensible not to train all potential users of the system but to have reference librarians help them. The console and the system, in effect, increase the reference librarian's memory by several orders of magnitude.

After that, there are catalogs of other libraries—especially specialized libraries—abstracts, and before long, I hope we can get some tables of contents or at least photographs in the graphic files, and then, ultimately, full text in digital form. Now at this point I'm really getting into the future, and I won't discuss that at all except to say that the system, as we see it, could go in this direction as we learn what to do with materials of that sort.

The Dynamic File Concept

In developing these files one thing that I talked about in my paper (I am not going to elaborate on it here), which I feel is the big value of this automatic system, is the concept that files and catalogs are not static things. I've mentioned a whole spectrum of files that have different kinds of material in them, and a search of even mild sophistication really should go from one file to another. Through a console the human being is asked to guide the search, but nevertheless a great deal can be done automatically if these files are set up in a dynamic way. This is a fundamental principle based on some ideas Turing had about 30 years ago. He's really the father of all computers and computing principles; he's been lost sight of. The reason I'm bringing him up here is that we have a new field, not numerical calculation, but nonnumerical processing, and it will pay us to go back to some of the fundamental ideas about data processing which are laid out in Turing's papers. And one way of interpreting what he said is just what I'm saying, that we can make these tables or files dynamic. Whenever you look anything up in one table, it always refers you to another one, even if it's a human being. We can do a great deal along these lines to create a very dynamic system.

I have gone a little bit into the future today on the principle that in designing a system there's a great deal of feedback; that is, when you come to the end of the line, you suddenly find out there are certain things that you should have done in the beginning. As a simple example, we talked about input to a system with all the different fonts and the output from the system with the limited fonts; well, if you are going to have a number of different characters in the output, obviously you have to make preparations for putting that symbolism in the input. Now this is obvious to everyone here, but this kind of feedback is not so obvious in many other ways, and I'd like to see as much feedback as possible put into the design of the system, based on what its future capabilities can be.

In summary, I'd just like to say that we survey team members who are here have been very interested in the discussion sessions, and I'm pleased to report that I don't think I heard anything new. This just means that at any rate we have thought about these things in the past, and I hope we've incorporated them properly in our study. Except for one thing that at least struck me as new: the system that we have been talking about, which is really just a large memory with a console for rapid access, is nothing more than a teaching machine. I think this is fine because what are libraries for but to teach us and make us find out new things?

Conclusion

Now just two things in conclusion: What is needed after this conference? The LC study group will break up when it finishes its report, although hopefully it might be replaced by some other group. We acted as a clearinghouse for different kinds of questions, but you can't look to us for all the answers. We certainly hope that all libraries who are interested, and that's a great many of them, will continue to expand their experiments and research into how libraries are used.

On the other side of the fence—if there's still a fence—there's the data processing industry, and their problem is that, to try to make it as simple as possible, they are trying to teach people the alphabet with an abacus. It can be done, but it's a very hard way of doing it. As soon as they can afford to develop machines for nonnumerical—that is, lexical—processing, the sooner we'll have equipment that's adapted to this problem that we've been discussing. For example, I suppose a hundred times in the last few days I've heard the word "tapes." Well, 2,000 years ago librarians got rid of tapes or scrolls; they invented the book. This is a great invention, but the computer people have hardly heard of it. Sure, they have disks, but they don't have any way of turning the page. So there's a lot to be done on the computer side to make equipment suitable for this vast field of application; we should not just try to adapt the present machines.

General Discussion

HENKLE: I've never been to a convention or a meeting in all my life when I've waited so long to say anything. I feel compelled to make a few remarks, if for no other reason than to try to reflect what I suspect is a kind of least common denominator of what goes on in the reactions and thoughts of quite a number of people here. Certainly I don't have the feeling that I've made any contribution to this session; it has made a great deal of contribution to me.

I would like to thank King. I would also like to thank Patrick, who is the epitome of those individuals who are impatient with us. There is some justification for his impatience, but I think that much of the reference that's been made at this meeting to two groups is a fictitious difference. There aren't really two groups here; there's one group because we have just one problem. It is not a problem that's new. Our big problem and our very common problem is that libraries aren't just the problem of librarians; they're the problem of scientists, because if libraries are the memory of our scientific culture, then the scientists have just as much stake in what's in them, just as much stake in how efficiently they're managed, and just as much stake in solving the problems of the librarians as any librarian in the room.

I would like to see the team, which Gilbert King has said is now breaking up because its immediate assignment is over, continued. I would like to see it include not only the particular half dozen people involved in the LC study but expanded to include the people who have made some very positive contributions to this meeting. I hope that it won't be very long until there'll be another conference just like this one, with this distinction: that the people who planned this conference review all of us here and pick out those who have demonstrated they can make some immediate contribution. These might convene 6 months from now so we don't let loose of these problems. They should spend even more time in preliminary preparation and come in not with documents designed to serve as a basis for discussion but with proposals designed to be evaluated for action. We have been talking about some of these problems for 10 years. I don't think we can wait another 10 years for the solutions. Now the one thing that I got out of this conference is that I don't think we have to.

GULL: Mohrhardt's remarks this morning encouraged me to think that the library side of this meeting was moving more rapidly than perhaps I had appreciated. King's initial remarks suggested to me that perhaps some of the attributes that I attributed to librarians might belong to the machine people as well.

I would like to observe that insofar as this conference has acquainted an outstanding and select group of librarians with aspects of our present day technology, which probably can be applied to the major problems and activities of large research libraries, it certainly has been successful. Insofar as it has required a select and competent group of technologists to look at library problems and to reduce them to facts, numbers, theories, and models, if you will, which can be used to relate these library problems to the available technology, it has also been a success. But this conference has been frustrating for me, and I think it has been to some others, because too many of the librarians present have been eager to demonstrate their newly acquired engineering talents, just as too many of the technical men have been willing to demonstrate their quick understanding of librarianship and have postulated neat solutions which ignore fully half of the significant considerations behind the problems to be solved. I suggest that, in this situation, the librarians should leave the engineering to the engineers, and the engineers should leave librarianship to the librarians, for together they have enough to do to cooperate with the systems people who have more than their work cut out for themselves in designing and implementing one or more workable, mechanized, and possibly automated, library systems.

This conference has clearly established, it seems to me, by virtue of the national representation working at the national level that none of these groups can accomplish our emerging objectives by working alone. This means, I believe, that the effect of this demonstration can be foreseen in relation to the forthcoming LC automation report.

Will the Library be able to evaluate this report alone or with the library profession? I think that the last few days indicate very strongly that this is not the case. Perhaps the most reasonable action then which can be expected is that nothing can be expected to be accomplished in the immediate future; I certainly hope this will not be the case. But the library problems which we are facing here are of such magnitude and of such importance to our lives that we can't entrust the evaluation of this report to the librarians alone. Another group must be formed for the evaluation of the report, and it should include at least, and obviously, selection from among librarians, engineers, information specialists, the users, and the systems people.

What can the library profession be doing while such an evaluation is, we hope, carried on? Certainly it can investigate the question which has been brought up already: Is automation necessary and desirable in libraries? If the answer is yes, and I hardly think that any other conclusion will be found, then the profession needs to heed the recommendations of the few who have spoken in the last couple of days—that the librarians must establish their goals and specify their requirements so that the people who can assist them can design and implement the workable solutions. It's been expressed on one or two occasions that this task is going to be particularly difficult for the members of a profession who are fully aware right now that they are not cooperating as well as they know how to cooperate and yet have persisted in this attitude over many decades. Higher degrees of cooperation, compromise, and standardization are going to be necessary as librarians go into mechanization. The profession must work together to approach their problems on a national level, to demonstrate their needs, to seek and obtain support, even if this means, for example, common support of one or more very serious projects out of the individual library budgets. Librarians can only pursue their present diverse paths if they are willing to agree upon, adapt, and implement a common framework for some national system.

Vosper: I'd like to question one small, but I think significant, point of King's paper and one that's echoed in what Gull just said. I fully understand the strategic advantages of attacking first the literature of science and technology as one looks at the economics of government at the present time. I would only like to urge, even recognizing the strategic and economic significance of this chronological establishment of value, that it would probably be morally wrong in undertaking such a major attack on the intellectual needs of the country to persist in a hierarchical value that the Government is already questioning and that society is questioning. There is a premium on the needs to solve the problems of science and technology, but we are closer to solving those needs than we are to solving other needs. I think we really should face up clearly to the total library system in which science and technology are only a very small part.

King: I absolutely agree with you, but I think you have to have a specific plan. Now I think none of us feels more strongly than you do that there is a total picture, but the total picture is a big bite. And a lot of the time we find it's like biting a big apple; we've got to find some way where we can get a relatively small mouthful first. It may be that in some of these other areas there would be a better way of doing it.

Alexander: Following this philosophy, then why tackle the oversized job of the Library of Congress instead of taking a smaller bite of the library apple? Why not start with one that could falter for a few years after this major surgery and not cause a tremendous upheaval?

King: Well, the main answer is that we were asked to look at the Library of Congress. The second answer is that we did consider creating a new library right from scratch somewhere in the desert or on the moon.

Fussler: Given the implementation of the system that is described in King's paper and outlined this morning, it seems to me the benefits to the technical processing operations of libraries are vividly evident in most respects. I would like to ask him to comment, however, on the problem that seems inherent in at least the initial years of operation of a system of this kind, with respect to the probability of its presenting to the reader an increasing amount of unevaluated information. I think one can envisage the reader-console dialogue reducing to some extent the bulk of the presented information with certain kinds of criteria, but the system itself is designed to add more information than the user would now get. In many typical operating situations, this is exactly what the reader doesn't want. He is not interested in an exhaustive bibliography or an extensive search, and it is one of the reasons that readers, so it seems to me, don't use libraries now. Instead they ask someone who knows; this is a complex, useful, and rather economical filtering process; it sometimes is quite effective. However, for an automated operation of this character, it seems to me there may be some real risks here, after the initial glamour of the console wears off a little bit. Its use, if qualitative evaluations or evaluations with respect to the reader's immediate criteria are not rather easily accessible, becomes frustrating too.

KING: We do have some very specific ideas about what this console should look like, and let's assume it has all the mechanical features we want. Now, how do I prevent myself from looking at too much? Well, I'm going to start off with some rather vague unsubstantiated things, because at first we only have a very simple card in there. It will be easier to turn these cards and, therefore, I think you can be more selective in not just stopping as soon as you've found something, but you can try and pick the right kind of author and the right kind of date. By having a full memory, you can have more of the information available, which of course is true of the standard LC card but not of all cards, and information like the publisher, to me anyway, is important in determining whether I really want to read that book or not. I can be more selective the more information of that type I see.

But I think that the most immediate way of improving the selectivity, which sounds like a paradox, is by having larger files. There's no question but that technology can supply all the memory you want at an incremental cost that's negligible; this means you can have much bigger files. Now you are trying to keep away from this; you don't have the space and it becomes cumbersome. But if you have a console, you can have as big a file as you like and the user doesn't know it. So the first thing is very primitive. We'll have a lot more see-also references and added entries of various sorts (now the tendency is to eliminate them), so that you can look through more of the catalog before you decide exactly what you want. To go a bit into the future (and I think this is quite a way away), the same kind of system could tolerate images of tables of contents. Maybe this is something we could start soon for some kinds of books, so that before you even ask for the book, you can look through the table of contents. I'm sure this would reject a lot of things, or it would make you selective.

Another thing that I think would be helpful would be to get good articles from encyclopedias into graphic files, so that when we want to start a library search, we could get a tutorial display from an encyclopedia for those people who wanted it, to give them a feel of what to ask for. We have to try to design the system to use every trick we can to be more selective. Is that a reasonable answer or is it just a promise?

FUSSLER: It's a reasonable answer, but I think the problem is a very difficult one to solve. I'm not sure that your answer disposes of it.

KING: I agree with you. We haven't gone into it sufficiently yet.

LIBBY: I'm going to make an attempt also to answer this question. I contend that by the use, in a well-designed system, of a process key, such as Don Swanson earlier mentioned, a query can be entered at the beginning of the search in terms that the person wants. Then the accumulated experience of reference librarians over years can be entered into a mechanized system to lead the user along a search path that can be tutorial, can introduce him to a subject, and so forth. You see this in printed form in the *Encyclopaedia Britannica;* the salesman makes a big issue of the fact that you don't have to wander indiscriminately through the pages of this encyclopedia; he has a little booklet that tells you to start on page so-and-so. I would contend that the automated system has a greater possibility of accumulating reference know-how and knowledge over a period of time and maintaining it for posterity than the human system now has.

WOOSTER: There are a lot of people who are some day going to have to mechanize their libraries. My own recommendation is that if you've never done anything in this area before rent a keypunch for $10 a month, experiment with card sorting equipment, get a feel of what's involved. Now it has been said here that if you can't do the whole job at once, it's not really worth starting at all. What is your feeling on these two approaches: get your feet wet by starting on a small scale, or by moving in one fell swoop?

KING: Under the system point of view you have to look at everything, integrate it all. The job of integrating any library to make it completely automatic is a very big job at the present time. The problem is how we can get our feet wet, how we can get started, without having to automate each thing, let alone a complete catalog, even in a medium-sized library. I feel strongly that we won't learn anything from the keypunch. I'll tell you what the result of this keypunch is—the reference librarians can do it better. In fact, just to pursue this argument, I have a rule of thumb that any human being can be very thoroughly acquainted with 10,000 documents; there are a lot of examples to substantiate this. So to prove that automation is better, you have to have many times more than 10,000 documents in your experiment.

HEILPRIN: It seems to me the basic problem of information systems in libraries is that the amount of information, the universe of discourse, is increasing, whereas the channel through which we take in the information remains constant. We take in 50 bits per second, approximately, and our problem is how to get at something in an increasing store through a rate-limited channel. There is only one human invention that has been made that can solve this, basically, and that is a system of dividing up this store by something we call classification, which allows us to get access to that part in which we're interested. So to me the basic solution as proposed by these matching sentences, transformations, and so forth, always has to come down to this: we have to set up a system of associations that are constantly increasing in complexity. That is, our programing will have to change constantly in such a way that we're not using all search paths, but as new things come in, they will be added to the existing search paths and provide sharper and sharper classes so that our finite rate of looking will still be satisfied. We get, in other words, a smaller and smaller mesh in our association net. This will have to be the nature of the ultimate program if a console, which is nothing but a switching device for association, is to be effective.

EDMUNDSON: I think it was very useful to hear a characterization of an automated library as a teaching machine. I'd like to point out that the automated library can also be regarded as a learning machine, in that the traces through the search path can be recorded by means of the console so that previously successful search strategies can be recorded and used time and time again. This is most clearly illustrated at present by the compilation of a special bibliography merely to have it erased from the system. The user, in this case perhaps the librarian himself, can specify that he has successfully answered this question. The answer can be recorded and put in a special part of the memory so that it can be produced without a complete retracing when that question is asked again.

HEILIGER: It is common practice now in reference departments to keep a record of the trail followed in answering difficult questions. But what I want to comment on was this matter of the mathematical model. We had a mathematical model made for our computer-based university library system; it was made by systems experts after considerable orientation in the library operation and considerable interaction with the entire library staff; I think this ought to be made clear. If any of you want to see a mathematical model for a full system for a university library, it has been published in the book that we issued last summer (item 67, p. 139).

KING: There are different kinds of mathematical models; you are talking about one kind. I was talking about a slightly different kind of model, trying to describe the actual lookup functions, how you find the things in the library, not just the flow of the information.

BERUL: As Heiliger said, some of these things are currently done to some extent with manual systems; for example, when a special bibliography has been prepared manually, one copy is generally kept for future use. The problem is how do we know that this bibliography has been prepared. Librarians keep records in the catalog, for example, of the fact that a special bibliography on a certain subject is some place on the shelf. The content of the special bibliography would probably be stored graphically, either in a hard copy or in microform. The search trail itself (that is, the trail leading to this special bibliography) might be stored in your machine for the man-machine interface but not the complete bibliography. It would be senseless to duplicate these in a separate file and store them digitally.

ANGELL: I share Fussler's concern for getting too much information into the store. We should be on guard against some of the implications of the things that are said about what lies ahead of us in automated possibilities. I refer again to this matter of depth indexing. If we mean by depth indexing what King said about getting an encyclopedia article into the store, Amen! (*Britannica* still has one of the best monographs on the Crusades, I am told.) This we need, but a lot of what is said about depth indexing (not at this conference) implies that if a person comes into the library and asks for the boiling point of water, we will press a button and will not only answer his question, but record every place in the library where that question is answered. Let us be on guard against this! I think we do not need to get every fact of nature and experience under pushbutton control; what we do need under pushbutton control is knowledge that is forming, that is nascent. This is what we need to be able to get at quickly and completely, because there is very little of it, and there's no selection of that.

In one word, what will keep us from having too much is, first, more discriminating indexing and secondly, the human nervous system because it has marvelous powers of association as well as marvelous powers of blocking.

ATCHISON: I have been sitting quietly learning from you librarians. I have been in the computer field for about 12 years now and I have heard similar discussions in other fields. For example, we have worked at our computer center with people from our State bridge department; they sat down with us, and now they are designing bridges with computers, doing all of their own work, and we never see them anymore. A similar thing happened in the electrical engineering field when we were asked to help design an electrical system for a new community; the electrical engineers are now off on their own. This has happened in one area after another. Now the point that is basic here is the matter of a system. Another basic point is that the different groups have to sit down together, and at our computer center we are doing that now with the library people.

Now, although I haven't spoken to the group before, I have been speaking to many of you individually and asking you the question: "How do you automate a library?" I've had many different answers, and I think this is the way it should be. Heiliger said, "Let's sit down and flowchart the whole operation." Someone else said, "Let's do our serials." Another stressed the accounting operations. There is not going to be one and only one approach. You librarians have a large system and a large problem, and you will have to work very hard on this. In many instances, rather than the whole system, I think you will have to begin with a small part, if for no other reason than it's a matter of education, and this has been our problem throughout.

MOHRHARDT: Time has determined that we have to close this session and the conference program. I think that Mr. Clapp, who represents one of the sponsors, should deliver a benediction.

CLAPP: This appearance is not programmed and will be brief. This is not the time to enter any evaluation of what we have done. This will occur through natural processes in all of us in the ensuing days, months, and even years. I would like to thank the writers of the papers on behalf of the conference as a whole for their time and efforts toward making this meeting as worthwhile as it seems to me to have been. Thanks are also due to the discussion leaders and to you the participants. And now having said this, all we can do is pat ourselves on the backs for having participated in a very useful exercise.

APPENDIXES

APPENDIX I

Biographical Data on Conference Program Participants*

SAMUEL N. ALEXANDER, a graduate of the University of Oklahoma (B.S.) and M.I.T. (M.S.), is chief of the Data Processing Systems Division of the National Bureau of Standards, where he directs programs concerned with Government uses of automation techniques in data processing, information storage and retrieval, and automatic instrumentation and dynamic control systems. He formerly served as chief of the Electronic Computers Laboratory of the National Bureau of Standards, where he was involved in the development of digital computer technology. He previously was a senior engineer with Bendix Aviation Corp., physicist with the Department of the Navy, and an engineer with the Simplex Wire and Cable Co.

Mr. Alexander is a member of many professional and technical groups, including a task force of the Federal Council on Science and Technology, the National Academy of Sciences, the Advisory Committee on Computers in Research for the National Institutes of Health, and the Atomic Energy Commission's Computer Advisory Committee. He has served as a technical consultant to a number of U.S. Government agencies and to the Governments of Sweden and India on automatic data processing applications and technology.

JOSEPH BECKER has degrees in aeronautical engineering from the Brooklyn Polytechnic Institute and library science from Catholic University. He served as a research fellow at the Western Data Processing Center at the University of California at Los Angeles, was a librarian at the New York Public Library, and served as coordinator for the American Library Association in connection with the Library 21 Exhibit at the Seattle World's Fair. Mr. Becker, a member of the American Library Association and the Association for Computing Machinery, recently was coauthor of a textbook entitled *Information Storage and Retrieval*.

LAWRENCE H. BERUL, a graduate of Drexel Institute of Technology (B.S.) and George Washington University School of Law (Juris doctor), is a senior systems engineer and director of the Washington office of the Information Dynamics Corp. In his present position he is the principal investigator under contract with the National Bureau of Standards to prepare a state-of-the-art report on output printing systems for producing abstracting and indexing journals. Previous assignments have included systems design of a mechanized photocomposition system for the NASA announcement journal *Scientific and Technical Aerospace Reports*. Mr. Berul formerly was staff attorney at C-E-I-R, Inc., where he concentrated on the field of legal information retrieval. From 1958 to 1961, he worked at the U.S. Patent Office as a patent examiner and as a systems and management analyst responsible for the development and review of data processing applications within the Patent Office.

Mr. Berul is a member of the Association for Computing Machinery and the American Bar Association and the author of several technical publications.

*This list is based on information available at the time of the conference. Changes in affiliation announced later than the conference are not reflected in the biographies.

DONALD V. BLACK, a graduate of the University of California at Berkeley (A.B., B.L.S.), is now associated with the Planning Research Corp. at Los Angeles. He was director of the Library Operations Survey, University of California at Los Angeles, as well as physics librarian and engineering reference librarian at that institution. He participated in several experiments concerned with the possibilities of automating the subject analysis of printed materials and the retrieval of such materials, directed a systems study of some of the functions of a major university library, and worked on large military projects involving the handling of linguistic data by computers.

HENRY J. DUBESTER, chief of the General Reference and Bibliography Division, Library of Congress, also served as coordinator for the survey group studying the feasibility of automating the Library. A graduate of the College of the City of New York (B.S.S.) and Columbia University (M.A.), his service with the Library of Congress was interrupted by 3 years of service in the Army Air Corps. Following World War II he became chief of the Census Library Project in the Library of Congress where he was concerned with the compilation of bibliographies and provision of reference service in demographic statistics. Mr. Dubester is a member of a number of professional groups, including the International Committee for Social Sciences Documentation and the U.S. National Committee for the International Federation of Documentation.

J. W. EMLING, executive director of the Transmission Systems Engineering Division of Bell Telephone Laboratories, Inc., is responsible for the systems engineering aspects of all types of transmission systems and for human factors research on communication systems. He has worked on many systems engineering studies in the fields of engineering economy, voice-frequency transmission, rural carrier, radio, television, and the transatlantic telephone cable system. In World War II he was engaged in the study of underwater acoustics.

Mr. Emling, a graduate of the University of Pennsylvania (B.S. in E.E.), is a member of the Acoustical Society of America, the Institute of Electrical and Electronics Engineers, and the American Association for Advancement of Science.

JAMES R. HARRIS, a graduate of the University of Richmond (B.S.) and Polytechnic Institute of Brooklyn (M.S.), is director of the Data Transmission Systems Engineering Center in Bell Telephone Laboratories, Inc. His technical experience includes the development of data switching and transmission channels, the development of high-speed computers for the U.S. Air Force, and the development of airborne communication and navigation equipment. Mr. Harris is a member of the Administrative Committee of the Computer Group, Institute of Electrical and Electronics Engineers.

GILBERT W. KING, vice-president and director of research, Itek Corp., heads research and development work in advance information technology in five laboratories specializing in optics, photography, chemistry, electronics, and information sciences. Dr. King has been associated with International Telemeter, where he was responsible for technical developments in connection with a photostore for lexical storage, and with International Business Machines Corp., where he initially had direction of programs in automatic language translation and information retrieval and more recently had responsibility for directing all research in the IBM laboratories.

Dr. King is a member of the President's Science Advisory Committee Panel on Problems of Scientific Information, chairman of the Library of Congress automation survey team, and is a member of the U.S. Air Force Scientific Advisory Board. He has served on the Visiting Committee for the M.I.T. Corporation and on the Air Force's Beacon Hill Intelligence Study, and has been a consultant to the U.S. Navy and to the Institute for Defense Analysis. Since 1956 he has worked closely with the U.S. Air Force on machine translation; during World War II he was a member of the Office of Scientific Research and Development where he was concerned with the use of data processing machines for the analysis of scientific data.

Dr. King has been associated with the California Institute of Technology, Harvard, Princeton, and Yale Universities, and the Massachusetts Institute of Technology. His research interests were in quantum and statistical mechanics and information theory applied to infrared spectroscopy.

RICHARD LIBBY, a graduate of the University of Massachusetts (B.S.), has done advanced study in physics, electrical engineering, and chemistry at the University of Maryland, George Washington University, and Syracuse University. He has served as consultant and technical staff member at the IBM Research Center and in a variety of administrative positions at the Rome Air Development Center (USAF), where he organized and directed developmental programs in ground-based electronic countermeasures, intelligence data handling, radio communications, and electromagnetic interference reduction. Mr. Libby was chairman of the Air Research and Development Command's Working Group on Intelligence and Reconnaissance, and in earlier service with the Naval Research Laboratory, he concentrated on the design and development of radio direction finders and other radio detection and countermeasure devices. Mr. Libby is a member of the survey team studying the feasibility of automation in the Library of Congress.

HARVEY J. MCMAINS, administrator for Data Communications Planning, American Telephone and Telegraph Co., is responsible for coordination of marketing activities involved in the development of data services. Educated at the universities of Georgia, Texas, Oklahoma, and Notre Dame, he holds undergraduate degrees in physics and mathematics and a master's in physics. Mr. McMains is a registered professional engineer and is the author of numerous articles in the fields of physics, mathematics, and data communications. He has experience as chief engineer of the Southwestern Bell Telephone Co. and at Bell Telephone Laboratories, Inc., where he worked on the development of transistors and similar conductor devices.

FOSTER MOHRHARDT, director of the National Agricultural Library, holds degrees from Michigan State University, Columbia University, University of Munich, and the University of Michigan. He has been associated with Brookhaven National Laboratory, the School of Library Service at Columbia University, and the Library Division of the U.S. Veterans Administration. As president of the International Association of Agricultural Librarians and Documentalists, Mr. Mohrhardt has participated in many international conferences. Long interested in the problems of science information and information control generally, Mr. Mohrhardt is currently a member of the Committee on Science Information of the Federal Council for Science and Technology and the Science Information Council of the National Science Foundation. He has been honored by the American Association for the Advancement of Science and the Institute of Information Sciences in London.

ROBERT L. PATRICK, a freelance computer specialist and consultant, is a graduate of the University of Nevada (BSME). He has held positions as consultant with the Computer Sciences Corp. at Los Angeles, as deputy director of the Computer Services Division of C-E-I-R, Inc., in Washington, and as aerophysics engineer with the Convair Division, General Dynamics Corp. in Fort Worth. He was a first lieutenant with the U.S. Air Force. His technical experience has included the development of aircraft structural design computations, gas turbine simulation procedures, fire control optimization codes, the design of a monitor system for an automatic computer operation, research in tape-controlled diesinking processes, and design of a data processing compiler.

FRANK B. [BRAD] ROGERS, Director of the National Library of Medicine at the time of the conerence, holds degrees from Yale University, Ohio State University's School of Medicine, and Columbia University's School of Library Service. A member of the U.S. Army Medical Corps until 1960, he is now a member of the Commissioned Corps of the U.S. Public Health Service. He is the current president of the Medical Library Association and was general chairman of the Second International Congress on Medical Librarianship. Dr. Rogers retired from the National Library of Medicine effective August 31, 1963, to become librarian and professor of medical bibliography at the University of Colorado Medical Center in Denver.

F. CLAYTON ROSE received a B.S. in engineering from the U.S. Naval Academy (1955) and has done graduate work in law at William and Mary and at Georgetown University Law Center. While in the Navy he had advanced work in communications and management in various service institutes. In his present position as data processing systems analyst in the Data Processing Systems Division, National Bureau of Standards, Mr. Rose is concerned with problem analysis, current awareness, evaluation, and abstracting in the fields of graphics, machine readability, character recognition, and the general field of machine applications for information storage, search, and retrieval. Mr. Rose previously was an engineering consultant at the Prevention of Deterioration Center, National Academy of Sciences, and supply officer for the U.S. Naval Communications Station, Washington. He is a member of the American Management Association, has assisted in the preparation of several technical reports, and has written a book on cipher analysis.

DAVID E. SPARKS, a graduate of Swarthmore College (A.B.) and Catholic University of America (M.A.), has a diploma in photographic technology from the Rochester Institute of Technology and was an exchange student at the University of Paris. He specialized in languages, library science, and information systems engineering and development. Mr. Sparks is currently a systems engineer with the Information Dynamics Corp., where he is the project manager of a study concerned with the development of systems concepts in the national dissemination of scientific information. Formerly librarian and information engineer with Itek Corp., Mr. Sparks has had experience in the design and development of information handling systems, including the application of information processing equipment to a variety of library operations, the analysis of information flow in a large military library, and the development of special techniques and equipment for mechanical manipulation of library catalog card data. Before going to Itek, he was associated with the General Electric Co. and the University of Vermont. A member of several professional library and technical associations, Mr. Sparks has published several technical studies concerned with the application of mechanical equipment and processes to certain phases of library work.

DON R. SWANSON, dean of the Graduate Library School at the University of Chicago, is a physicist with degrees from the California Institute of Technology (B.A.), Rice Institute (M.A.), and the University of California at Berkeley (PH.D.). After experience in the Radiation Laboratory of the University of California and in research computer application at Hughes Aircraft Corp., Dr. Swanson became manager of the Synthetic Intelligence Division at Thompson Ramo Wooldridge, Inc., where his interests centered on problems of automatic indexing, computer application in intelligence analysis, and machine translation. He is a member of the survey team that has been studying the problems connected with mechanization of the operations of the Library of Congress and has written many technical publications.

MORTIMER TAUBE founded Documentation, Inc., in 1951 and is currently chairman of its board of directors. He is known for his contributions in the field of information theory and has served on numerous committees concerned with national and international documentation. He is currently adjunct professor at the Columbia University School of Library Service and has lectured in the graduate schools of Columbia and the University of Chicago. A graduate of the University of Chicago, Dr. Taube studied philosophy at Harvard University and received his PH.D. degree from the University of California. His library experience includes service at Mills College, Rutgers University, Duke University, and the Library of Congress. Dr. Taube has served for some years as American representative on the Documentation Committee of UNESCO and is the author of many technical publications. He has also directed a wide range of data and information studies for industry and government.

DAVID P. WAITE, president of Information Dynamics Corp., is a graduate of the University of Pennsylvania (B.S.). He is the designer of the microform system used by the National Aeronautics and Space Administration for technical report dissemination. He has undertaken engineering and cost analysis studies for large serial collections in libraries, has participated in the study of the design of projection displays for teaching machines, and has responsibility for the design of a unitized system for handling weather satellite cloud photography by meteorological researchers.

Formerly with Itek Corp., General Electric Co., and the Bartol Research Corp., Mr. Waite has analyzed insurance operations, technical reference and library systems, management of medical records, and military intelligence and reconnaissance systems, and directed the engineering team that developed steam generator controls for , uclear submarines and surface ships for the U.S. Navy. Mr. Waite has many technical publications to his credit and is a member of several professional associations.

ALBERT WARHEIT, a graduate of the University of Michigan (A.B., M.A., PH.D.) with degrees in library science and linguistics, studied also at the University of Zurich, Switzerland. Long interested in problems of information storage and retrieval, he has published papers and participated in technical symposia. Currently associated with the information retrieval program in IBM's Advanced Systems Development Division, he is concerned with the development of indexing systems and document storage and data retrieval systems. Previous experience includes assignments with General Motors Corp. and the Atomic Energy Commission, where as chief librarian he organized an abstracting service, had responsibility for all AEC library services to laboratories and contractors, and established the worldwide library depository system of AEC publications. There he also installed the first punched card accounting system for the control and inventory of classified documents and mechanized the compilation of abstract journal indexes.

APPENDIX II

List of Conference Participants*

Burton W. Adkinson, Head, Office of Science Information Service, National Science Foundation

Samuel N. Alexander, Chief, Data Processing Systems Division, National Bureau of Standards

Richard S. Angell, Chief, Subject Cataloging Division, Library of Congress

George Arnovick, Staff Scientist, North American Aviation, Inc.

W. F. Atchison, Chief, Rich Electronic Computer Center, and Acting Director, School of Information Science, Georgia Institute of Technology

Roy P. Basler, Director, Reference Department, Library of Congress

Joseph Becker, Data Processing and Library Consultant

Richard Benedict, Assistant to the Director, University of Florida Libraries

John H. Berthel, Librarian, Johns Hopkins University Library

Lawrence H. Berul, Director, Washington Operations, Information Dynamics Corp.

Ralph Blasingame, Jr., State Librarian, Pennsylvania State Library

Raymond A. Bohling, Supervisor, Departmental Libraries, University of Minnesota Library

Harold Borko, Head, Information Retrieval and Linguistics Project, System Development Corp.

John Bowling, Supervisory Engineer, Electronics and Ordnance Division, AVCO Corp.

Roger P. Bristol, Departmental Librarian, University of Virginia Library

Margaret C. Brown, Chief, Processing Division, Free Library of Philadelphia

Mrs. Helen L. Brownson, Program Director for Documentation Research, Office of Science Information Service, National Science Foundation

Lawrence F. Buckland, President, Inforonics, Inc.

*Information about participants is taken from conference registration forms; changes in affiliation after the conference are not reflected in this list.

Thomas R. Buckman, Director of Libraries, University of Kansas Libraries

Robert E. Burton, Head, Science and Engineering Libraries, University of Michigan Library

Wayne R. Campbell, Chief Librarian, Scientific Library, U.S. Patent Office

Richard E. Chapin, Director, Michigan State University Library

Verner W. Clapp, President, Council on Library Resources, Inc.

William S. Dix, Librarian, Princeton University Library

Henry J. Dubester, Chief, General Reference and Bibliography Division, Library of Congress

H. P. Edmundson, Senior Staff, Thompson Ramo Wooldridge, Inc.

Ralph E. Ellsworth, Director, University of Colorado Library

J. W. Emling, Executive Director, Transmission Systems Engineering Division, Bell Telephone Laboratories, Inc.

Ralph T. Esterquest, Librarian, Harvard Medical Library

Edward J. Forbes, Electronic Printing Research Officer, Government Printing Office

Bernard M. Fry, Deputy Head, Office of Science Information Service, National Science Foundation

Herman Fussler, Director, University of Chicago Libraries

Alvin J. Goldwyn, Associate Director, Center for Documentation and Communication Research, Western Reserve University

Mandalay Grems, Staff Consultant for Systems Programming, UNIVAC Division, Sperry Rand Corp.

Hillis L. Griffin, Information Systems Librarian, Argonne National Laboratory

C. Dake Gull, Consulting Analyst, Information Systems Operation. General Electric Co.

Warren J. Haas, Associate Director, Columbia University Libraries.
Mrs. Elizabeth E. Hamer, Assistant Librarian, Library of Congress
Lillian A. Hamrick, Chief, Technical Information Division, Office of Technical Services, U.S. Department of Commerce
J. R. Harris, Director, Data Transmission Systems Engineering Center, Bell Telephone Laboratories, Inc.
Katharine G. Harris, Reference Services Director, Detroit Public Library
Robert M. Hayes, President, Advanced Information Systems, Inc.
Edward M. Heiliger, Librarian, Chicago Undergraduate Division, University of Illinois Library
Laurence B. Heilprin, Staff Physicist, Council on Library Resources, Inc.
James W. Henderson, Assistant to the Director, New York Public Library
Herman H. Henkle, Librarian, John Crerar Library
Mrs. Mary T. Howe, Librarian, Decatur [Illinois] Public Library
Mrs. Frances Jenkins, Professor, Graduate School of Library Science, University of Illinois
Harold Johnson, Vice President-Engineering, Photon, Inc.
Sidney Kaplan, Manager, Advanced Information Storage and Retrieval Systems, RCA Data Systems Center, Radio Corporation of America
David Kaser, Director, Joint University Libraries
J. Hilary Kelley, Technical Assistant, Office of Science and Technology, Executive Office of the President
Gilbert W. King, Vice President and Director of Research, Itek Corp.
Katharine Laich, Assistant City Librarian, Los Angeles Public Library
Mrs. Dorothy Levy, Cataloger, Catalog Department, Drexel Institute of Technology Library
Richard L. Libby, Director, Westchester Laboratory, Itek Corp.
Richard H. Logsdon, Director of Libraries, Columbia University Libraries
Frank A. Lundy, Director of University Libraries, University of Nebraska Libraries
Mrs. Barbara Evans Markuson, Assistant to the Information Systems Specialist, Library of Congress
Stephen A. McCarthy, Director of Libraries, Cornell University Libraries
Edward M. McCormick, Senior Research Analyst, Office of Science Information Service, National Science Foundation
Marvin W. McFarland, Acting Chief, Science and Technology Division, Library of Congress
Harvey J. McMains, Administrator, Data Communications Planning, American Telephone & Telegraph Co.
Robert A. Miller, Director, Indiana University Library
Thomas L. Minder, Supervisor, Library Research and Development, Pennsylvania State University
Foster E. Mohrhardt, Director, National Agricultural Library
Edward B. Montgomery, Research Consultant, Syracuse University
John Moriarty, Director, Purdue University Libraries
Mrs. Marlene Morrisey, Executive Assistant to the Librarian, Library of Congress
L. Quincy Mumford, Librarian of Congress
Gerhard B. Naeseth, Associate Director, University of Wisconsin Libraries
John A. Neal, North American Aviation, Inc.
Jerrold Orne, University Librarian, University of North Carolina Library
John Henry Ottemiller, Associate University Librarian, Yale University Library
Howard E. Page, Head, Office of Institutional Programs, National Science Foundation
Robert L. Patrick, Computer Specialist, Planning Research Corp.
Paul Poindron, Conservateur en chef, Direction des Bibliothèques de France, Ministère de l'Education Nationale, representing the Bibliothèque Nationale
Frazer G. Poole, Director, Library Technology Project, American Library Association
William B. Quirk, Manager, Data Communications Planning, American Telephone & Telegraph Co.
Gordon E. Randall, Librarian, Thomas J. Watson Research Center Library, representing the Special Libraries Association

Mrs. Phyllis A. Richmond, Supervisor, River Campus Science Libraries, University of Rochester Library

Joseph H. Roe, Jr., Head, Reference Department, National Library of Medicine

Frank B. [Brad] Rogers, Director, National Library of Medicine

Rutherford D. Rogers, Deputy Librarian of Congress

F. Clayton Rose, Data Processing Applications Analyst, Research Information Center, National Bureau of Standards

Frank L. Schick, Assistant Director, Library Services Branch, U.S. Office of Education

John Sherrod, Chief, Information Services and Systems Branch, Division of Technical Information, U.S. Atomic Energy Commission

James E. Skipper, Executive Secretary, Association of Research Libraries

Richard L. Snyder, Associate Director, Massachusetts Institute of Technology Libraries

David E. Sparks, Library Systems Engineer, Information Dynamics Corp.

Mary Elizabeth Stevens, Supervisory Operations Research Analyst, Information Technology Division, National Bureau of Standards

Don R. Swanson, Dean, Graduate Library School, University of Chicago

Robert L. Talmadge, Director, Tulane University Library

Mortimer Taube, Chairman of the Board, Documentation Inc.

Robert S. Taylor, Director, Center for the Information Sciences, Lehigh University

Frederick R. Theriault, Department of Defense

George Vdovin, Head, Public Services Department, University of California, San Diego

Melvin J. Voigt, University Librarian, University of California, San Diego

Robert Vosper, University Librarian, University of California, Los Angeles

David P. Waite, President, Information Dynamics Corp.

I. A. Warheit, Senior Systems Analyst, Advanced Systems Development Division, International Business Machines Corp.

David C. Weber, Assistant Director, Stanford University Libraries

William J. Welsh, Associate Director, Administrative Department, Library of Congress

Gordon Williams, Director, Midwest Inter-Library Center

Harold Wooster, Director of Information Sciences, Office of Scientific Research, U.S. Air Force

☆ U.S. GOVERNMENT PRINTING OFFICE: 1967 O—241-308

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $2.75