tems files are organized and arranged in an endless variety of ways. Generally, the method used initially for organizing and arranging the data prior to conversion to a computerized system is also the method selected for the new system. Thus, computerized census records are organized and arranged on a geographical basis much as they were before the advent of the computer. Personnel data banks are usually organized by the name or identification number of individual employees or job applicants. However, the computer offers one distinct advantage not normally possible or practical in conventional systems—the capability of organizing and arranging the same data in a variety of other ways. For example, personnel data can, in addition to the basic arrangement, be organized on the basis of organizational assignment, position classification series, years of service, etc., for direct searching or preparation of special listings.

Case files (files organized by the names or identifying numbers of people, places, or things) represent approximately 85 percent of the folderized records of the Federal Government. These files contain a wealth of data, but when stored in conventional systems the data is buried so deep in the file that it receives only limited use. By converting the data in these files to computerized systems, it becomes possible to readily select, extract, compare, and manipulate the data in an endless variety of ways to meet day-to-day operational requirements, to provide statistical data for management decisions, and to satisfy unpredictable needs of the future.

The only serious disadvantage of computer data storage and retrieval systems at present is their cost. However, the cost picture is gradually changing due to reduction in computer input costs through the application of SDA techniques; larger and cheaper computer data storage devices; faster processing speeds; and faster, less costly methods and equipment for retrieving and producing the system output.

Tomorrow's records manager will more than likely discover that most of the data needed to satisfy his clientele will be available via the computer and that his conventional files will serve mainly as depositories for selected original documents having legal or archival value. Today's records managers should therefore survey every existing record series for the purpose of identifying those which at some future date will or should be converted to a computerized data base and then work with management in developing an orderly schedule for the conversion.

Other Machine Indexing and Retrieval Systems

While most of the microform equipment described in chapter III is designed primarily for storage of documents or data in miniaturized form, some also have the capability to conduct logic-type searches. These are as follows:

Motorized (mechanized) Roll Microfilm with Photo-optical Binary Code. Although retrieval speeds with this type of equipment are not nearly so fast as those that are possible with a computer, they permit the user to automatically retrieve information. The information is displayed in page size, usually on a viewing screen, or reproduced on a film or paper copy. However, data on the film cannot be moved from one location to another, nor rearranged or changed. (For further information, see chapter III.)

Microfilm Chip, Automated. This equipment has about the same capabilities as the system described immediately above. The use of the chips, however, does make it possible to insert and delete individual pages. (For further information, see chapter III.)

Aperture Card. (EAM punched card-microfilm). Systems of this type make it possible to mechanically sort, select, display, and copy printed or graphic information appearing on the film images displayed on the cards. However, as in the case of microfilm chip automated systems, the equipment is not well suited to personal searching by individual users. (For further information, see chapter III.)

Microform-Computer Combinations. Various types of microform equipment can be linked either directly or indirectly to a computer so that the computer can be used to conduct the searches and the microform device used to store and display the information or documents the user is seeking. (For further information, see chapter III.)

VI. HOW TO DECIDE IF A NEW SYSTEM IS NEEDED

The Preliminary Survey

This handbook gives considerable attention to finding the best system for storing and retrieving information. There will always be situations where the best system is the same system used in the past. Other situations will warrant the use of modern information retrieval methods and equipment.

Sometimes information retrieval studies are pursued for weeks or months, or a new system is installed, only to discover that a conventional system is all that is needed. The first question, therefore, that needs to be answered—and rather quickly—is "When do I use the old and when do I use the new?" This chapter describes a step-by-step procedure for making a preliminary survey to answer that question. It will help in deciding when conventional methods should be used and when it is worthwhile to spend the time and effort to make a detailed study of the possibilities of modern information retrieval methods and equipment.

Where to Look

The preliminary survey should not be limited to the major files, the library, or collections of reference materials. Rather, you should look anywhere there is a collection of information stashed away, regardless of the form in which it is stored. In this handbook, these files or other collections are referred to as "information facilities." Certainly, the size and frequency of use of the information facility are considerations, but they are less likely to rule out any system than they are to affect the type of system needed when weighed on the cost-benefits scale. Small units can sometimes justify relatively inexpensive and yet modern information retrieval systems. This is particularly true where there are many small information facilities containing information all or a substantial portion of which is the same.

For further clarification of the wide potential, consider any of the following situations:

Case-type records used to correlate or compare data relating to individual persons, places, or things, for such purposes as personnel selection and placement, selection of contractors for bidding, selection of equipment, and conducting special analyses.

Case-type records used for looking up and extracting discrete data such as names, addresses, amounts, dates, and other data needed for such purposes as answering correspondence, processing applications, and preparing reports.

Subject files and indexes relating to written text and used for obtaining any information that might aid in handling a current task or problem in connection with such activities as legal work, research, preparation of instructions, and management planning.

Reference collections containing such items as publications, technical reports, procedural manuals, directories, catalogs, and statistics used in day-to-day operations or research.

Files of graphic or pictorial material such as maps, photographs, slides, and engineering or architectural drawings in situations where the users are trying to find items having set characteristics or attributes.

Examining User Needs

Looking at all information facilities, of whatever description, is a practical and solid starting point. It is, however, at least equally important to examine the needs of the people who use the information.

Why is it important to look at both the infor-

SAMPLE PRELIMINARY SURVEY FORM FOR INFORMATION FACILITIES

Information Retrieval Preliminary Survey INFORMATION FACILITY	INSTRUCTIONS - Prepare one of these Reconnaissance Data Sheets for each file station, record collection, index file or other information facility at the installation being surveyed. Where reference is made to user man-hours, specify those spent by employees of the facility as well as any spent at the facility by personnel from other organizational units.	ANALYSTS NAME DATE

ORGANIZATION AND FACILITY

NAME AND ADDRESS OF ORGANIZATION OF JURISDICTION	INFORMATION CONTAINED IN RECORDS	RECORD PHYSICAL FORM	
	TOTAL NO. (Net) EMPLOYEES ASSIGNED TO OPERATE AND MAINTAIN FACILITY	BLDG & ROOM NO.	PHONE NO.
TITLE OF INFORMATION FACILITY	CURRENT VOLUME (No. of pages, if written info; No. of characters if precise data; No. of items, if graphics)	ANNUAL ADDITIONS	USEFUL LIFE OF INFORMATION

USAGE DATA - To estimate manhours spent at this information facility in looking up, searching, extracting, or correlating information or data - - - - including manhours spent by personnel assigned to the facility as well as time spent at the facility by personnel from other organizational units.

PRIMARY USERS (Organization and Unit)	JOB TITLE	TOTAL ANNUAL MANHOURS	TYPE OF RETRIEVAL ACTION

DESCRIBE TYPICAL LOOKUP, SEARCH, DATA EXTRACTION OR CORRELATION ACTIONS

DESCRIBE ANY INADEQUACIES, PROBLEMS, OR LIMITATIONS OF THIS INFORMATION FACILITY

REMARKS

mation facilities and the users' needs? Why is it not sufficient to stop with a look at the demands upon and limitations of the information facilities themselves? There are many reasons, but the following are particularly significant:

- Data gathered at the information facility or from the users alone would be incomplete and misleading; whereas gathering information from both serves to supplement and cross check the information furnished by the other.

- Personnel operating an information facility cannot always describe or interpret user needs accurately.

Users' statements must be weighed in the light of actual information facility experience:

- If the information facility receives moderate or heavy use, the users probably have a real need for information—perhaps for even more than they are now getting.

- If the facility receives only light use, the probability of an urgent users' need is suspect unless the facility is not readily accessible nor operated properly.

Fact-Gathering Forms

The person conducting the preliminary survey should, if possible, personnally collect the data relating to the information facilities and users' needs, in which case the data could be recorded directly on decision tables similar to those shown in figures 36 and 37. If, however, the information users and the personnel operating the information facilities will be requested to supply the data themselves, the use of forms similar to those shown in figures 34 and 35 is suggested.

Information Retrieval Preliminary Survey—Information Facility (Fig. 34). This form may be used for collecting data about the various file stations, manual or machine record files, publications, and any other collections of typed, handwritten, printed, or graphic material. The data appearing on these forms, together with the personal knowledge of the individuals who completed them, will later serve as the basis for preparing information facility decision tables.

Information Retrieval Preliminary Survey—User Needs (Fig. 35). This second form may be used to obtain a sampling of how much time the users are now spending in looking up, searching, extracting, or correlating information or data, and to identify any inadequacies, problems, or limitations of the present sources or methods. These completed forms will also be used later for preparation of decision tables.

Decision Tables

Two decision tables have been prepared to help show what conclusions may be reasonably drawn from any set of facts gathered. These tables require the answering of various "yes" or "no" questions about the facts. The patterns shown by the "yes" and "no" answers lead to certain predetermined conclusions shown on the forms. One table is for analyzing facts gathered about the information facility and the other relates to facts about users' needs. Blank copies of these two decision tables are included as Appendix "D." Figures 36 and 37 provide filled-in examples of the two tables.

Evaluating Information Retrieval System Potential—Information Facility. (Fig. 36).

This form contains spaces for entries of certain identification and usage data at the top. Then, under "Evaluation Factors," a "Y" or "N" should be entered under the "Yes" or "No" column for each factor, depending upon your findings. The resulting yes-or-no pattern in this column is the same as one of the columns under "Key." It is this pattern that identifies the conclusion appropriate for the particular set of facts being analyzed. The "yes" and "no" answers might be thought of as "votes" for or against a modern information retrieval system (except for No. 5 evaluation factor, which is reversed). But it is not merely a matter of counting up affirmative and negative answers, since some evaluation factors carry more weight than others. It is the exception rather than the rule that the decision for or against would be based on just one of these factors.

SAMPLE PRELIMINARY SURVEY FORM FOR USER NEEDS

Information Retrieval Preliminary Survey USER NEEDS	INSTRUCTIONS: Prepare one or more of these Reconnaissance Data Sheets, as needed, for each of the broad, similar types of information or data needed by the installations being surveyed. *EXAMPLES: Personnel data, legal precedents, accounting data, research information, etc.*	ANALYST'S NAME DATE
BROAD TYPE OF INFORMATION OR DATA (Short title)	BRIEF DESCRIPTION OF THE INFORMATION OR DATA	

USAGE DATA – To identify individual user groups and the manhours they spend in looking up, searching, extracting, or correlating this information or data.

ORGANIZATIONAL UNIT	USER'S JOB TITLE (Exclude personnel assigned to operate information facilities)	NUMBER OF USERS	USER'S PHYSICAL LOCATION	ANNUAL MANHOURS (All users)	PRIMARY SOURCES OF THIS INFORMATION (Name and location)

DESCRIBE TYPICAL LOOK UP, SEARCH, DATA EXTRACTION, OR CORRELATION ACTIONS

DESCRIBE ANY INADEQUACIES, PROBLEMS, OR LIMITATIONS INVOLVED IN THE PRESENT SOURCES OR METHODS EMPLOYED

REMARKS

Figure 35

These are some of the basic concepts involved in the following "Evaluation Factors":

Factor 1: "Annual additions equal or exceed." Modern information retrieval systems are normally designed to handle fairly large collections of information or data. The addition of 25,000 pages or 2,500 individual graphic items annually or the maintenance of one million characters of data that are constantly being updated may be considered the minimum volume requirement for a positive vote for modern information retrieval methods. It is possible to have less volume and still find some need for an information retrieval system, but the probabilities are less likely. A "no" vote, therefore, does not necessarily rule out the potential need for an information retrieval system.

Factor 2: "Information will be in continuous use for over 5 years and one man-year or more is being used for looking up, searching, extracting, or correlating information or data at this facility." Because information retrieval systems always create new and often considerable expense, particularly in the input phase, they are ordinarily not used for information or data of short term value. And unless coupled with at least 1 man-year of work in searching, etc., there may not be enough potential manpower savings to offset the cost of an information retrieval system. A "yes" answer here is another vote for information retrieval, but by no means a justification in itself.

Factor 3: "Information will be in continuous use for less than 5 years and two man-years or more are being used for looking up, searching, extracting, or correlating information or data at this facility." The extra expense of an information retrieval system might be justified even though the information or data were of shorter use value if there is a potential for saving two or more man-years of searching time. Evaluation factors 2 and 3 are mutually exclusive—in a given situation only one could apply. Also, of course, in some instances neither may apply. Also note, as explained in the second sentence under "Instructions" at the bottom of the form, that the man-hours include both those of the personnel assigned to operate the facility as well as to others who come to conduct searches at the facility.

Factor 4: "Time presently required for looking up, searching, etc., information or data at this facility is mainly attributable to limitations of conventional methods." A "yes" vote is used here only when it can be determined that the reasons it takes so much time to retrieve information are due to the inherent limitations of conventional methods, and that it should be possible to reduce retrieval man-hours by installing a modern information retrieval system.

The fact that extensive man-hours are being spent to obtain information need not mean that the conventional system is inefficient. It may simply be due to the heavy workload. (In some situations a conventional system can retrieve information faster and cheaper than a modern information retrieval system.)

To evaluate this factor properly, one must therefore clearly understand the inherent advantages and disadvantages or limitations of both conventional and nonconventional methods.

Factor 5: "The information maintained at this facility could be readily obtained from other source(s)." Be sure to note that a "yes" vote here is a vote against a modern information retrieval system. This factor is included in the decision table because other places where the same information is available are sometimes overlooked. Modern transmission methods and duplicating services may make it more practical to use another source instead of maintaining a duplicate facility. By pooling the resources used to maintain the duplicate or complementary information facilities, it may also be possible to install a modern information retrieval system.

There follows explanations for the five conclusions depicted in figure 36.

Conclusion A: "A modern information retrieval system seems a likely possibility." This means only that from your observation at the present time, you can conclude that there is a definite possibility it may be profitable to install a modern information system.

Conclusion B: "Likely that present or improved conventional methods will suffice." This means that you have eliminated any reasonable doubt as to the need for a modern information retrieval system.

Figure 35

SAMPLE FORM FOR EVALUATING INFORMATION RETRIEVAL SYSTEM POTENTIAL—INFORMATION FACILITY

Evaluating Information Retrieval System Potential

INFORMATION FACILITY

EVALUATOR'S NAME: GEORGE ADAMS

DATE: 9 - 3 - XX

ORGANIZATION AND FACILITY

NAME AND ADDRESS OF ORGANIZATION OF JURISDICTION: Administrative Div. Office of General Counsel Denver Regional Office

TYPE OF RECORDS: [X] FOLDERS [ ] CARDS [ ] OTHER (Specify)

CONTENTS OF RECORDS: Claims applications, legal opinions and decisions, court records, and corres.

TITLE OF INFORMATION FACILITY: Central Legal Files

NO. (Net) OF EMPLOYEES AT FACILITY: 3

BUILDING AND ROOM NUMBER: Rm 600 Admin. Bldg

PHONE NO: X39290

USAGE DATA (*Estimated manhours spent annually in looking up, searching, extracting or correlating information or data at this facility*)

PRIMARY USERS (Organization & Unit)	JOB TITLE	ANNUAL MANHOURS	PRIMARY USERS (Organization & Unit)	JOB TITLE	ANNUAL MANHOURS
Claims Div	attorney	1000	Gen. Law Div.	Law clerk	6000
Litigation Div.	attorney	1500	Law Library	Librarian	2000

EVALUATION FACTORS	YES OR NO	KEY									
		✓									
1. Annual Additions Equal or Exceed: (*Circle applicable letter, if any*) a. 25,000 *pages, if system is used mainly for storage of written information.* b. 1,000,000 *characters, if system is used for storage of precise data such as names, numbers, etc.* c. 2,500 *individual items, if system is used mainly for storage of graphic, pictorial, or other matter not covered above (Explain in remarks).*	Y	Y	Y	Y	Y	N	N	N	N	N	-
2. Information will be in continuous use for over 5 years and one man-year or more is being used for looking up, searching, extracting, or correlating information or data at this facility.	Y	Y	N	N	N	Y	Y	N	N	N	-
3. Information will be in continuous use for less than 5 years and two man-years or more are being used for looking up, searching, extracting, or correlating information or data at this facility.	N	N	Y	Y	N	N	N	Y	Y	N	-
4. Time presently required for looking up, searching, etc., information or data at this facility is mainly attributable to limitations of conventional methods.	Y	-	Y	N	-	Y	N	Y	N	-	-
5. The information maintained at this facility could be readily obtained from other source(s) (*Specify sources and locations under remarks*).	N	N	N	N	N	N	N	N	N	N	Y
CONCLUSIONS											
A. Modern information retrieval seems a likely possibility.		X	X								
B. Likely that present or improved conventional methods will suffice.				X	X		X		X	X	
C. Likely that present or improved conventional methods will suffice; HOWEVER, also consider modern information retrieval systems (*particularly those which use inexpensive equipment.*)						X		X			
D. Consider discontinuance of either this or other duplicate facility(ies), and if duplication is widespread, also consider possibility of a central information service or facility.											X
E. Other (*Specify and explain - use remarks if additional space is required*).											

REMARKS

INSTRUCTIONS - Prepare one of these Decision Tables for each file station, record collection, index file or other information facility at the installation being surveyed. Where reference is made to user manhours, specify those spent by employees of the facility as well as any spent at the facility by personnel from other organizational units. Answer "YES" or "NO" in the appropriate column opposite the Evaluation Factors to indicate the existing situation. Compare your overall findings with those in the columns under the KEY. (*A dash indicates that it makes no difference whether the answer to that evaluation factor is Yes or No.*) When you find a column that duplicates your answers, place a check mark at the top of the column (*preferably with a colored pencil*). Follow the appropriate column down into the Conclusions column and circle the appropriate X.

Figure 36

Conclusion C: "Likely that improved conventional methods will suffice; however, we should also consider modern information retrieval systems." This represents a "gray" area situation that you will probably not want to eliminate at this time.

Conclusion D: "Consider discontinuance of either this or other duplicate facility(ies), and if duplication is widespread, we should also consider the possibility of a central information service or facility." This is self-explanatory.

Conclusion E: "Other." This permits the person making the study to provide an alternate conclusion or to take exception to what would have been the normal conclusion due to factors not covered in the decision table: for example, if it were found that a major change in the functions, workload, or organizational structure were imminent.

Important: Note that the block at the top of figure 36, titled "No. (net) of Employees at Facility," refers to the net number of people (or man-hours) required for operating the facility, even though some situations may require only a small portion of the total staff for searching the files, the remainder being used to enter information into the system and keep it in proper condition. (This item should not be confused with the man-hour figures called for in evaluation factors 2 and 3.)

Evaluating Information Retrieval System Potential—User Needs (Fig. 37). This form is used and analyzed in the same manner as the information facility form in figure 36. These are the basic concepts involved in its evaluation factors.

Factor 1: "5 percent or more of users' total man-hours (minimum 1 man-year) are being spent in looking up, searching, extracting, or correlating information or data." The probability is that a modern information retrieval system will not be considered unless it can be justified economically. Hence, the more time that users spend in trying to get the information needed, the greater the possibility of saving their time and offsetting the cost of information retrieval systems. If the users spend less than 5 percent of their time in such efforts, it is unlikely that information retrieval can recover enough of the users' time to pay for the system.

Factor 2: "Current information facilities are inadequate for one or more of the following reasons." These represent disadvantages or deficiencies of conventional systems from the viewpoint of the users. Often these problems can be overcome through application of modern information retrieval methods. Factor 2 should be answered "yes" only when the problem is inherent in the conventional system employed, not when it is due to faulty design or operation. A "yes" vote here is therefore a vote for a modern information retrieval system.

Factor 3: "Much faster retrieval speed is needed than could ever be achieved under present or any other conventional method." If there is an overriding need for retrieval speed, there may be justification for a modern information retrieval system. This factor may be important enough to overrule negative responses to the other factors. Situations of this type often exist in intelligence work, defense systems, and sometimes in office areas, too.

Factor 4: "Time presently spent in searching, extracting, or correlating information or data is mainly attributable to limitations of conventional methods." The remarks for evaluation factor 4 for the information facility decision table also apply here. Further, a double check from the viewpoint of the user is necessary to make certain that the conventional system and equipment are the problem, rather than something else; for example, man-hours spent reading and examining documents after they have been retrieved, which is a common practice in some professions regardless of the retrieval system used. Therefore, to evaluate this factor properly the analyst needs to investigate present practices and procedures.

The explanations of the conclusions for this table are the same as the explanation offered for the table on information facilities, except for the omission of conclusion D, "Consider discontinuance of either this or other duplicate facility." This form also has an "Inconvenient Features" section at the bottom that is not part of the decision table itself but is supplementary in nature and is included for the following reasons:

- To make sure that the person making the study does not confuse mere inconvenience with inadequacy and thereby erroneously mistake the former for the latter in evaluation factor 2.

SAMPLE FORM FOR EVALUATING INFORMATION RETRIEVAL SYSTEM POTENTIAL—USER NEEDS

Evaluating Information Retrieval System Potential USER NEEDS	EVALUATOR'S NAME: HELEN DAVIS
	DATE: 9 - 3 - XX

BROAD TYPE OF INFORMATION: Personnel data (employee skills, education, experience, etc.)

ORGANIZATIONAL UNIT	USER'S JOB TITLES (*Exclude personnel assigned to operate information facilities*)	NUMBER	PHYSICAL LOCATION	ESTIMATED ANNUAL MANHOURS	PRIMARY SOURCES OF THIS INFORMATION
Personnel Div	Placement office	3	Rm 708 Main	1500	Central files
" "	Manpower sect.	6	Rm 723 "	2000	" "
Misc Supervisors	misc.	25+	misc locations	500	" "
Personnel Div	Training off.	3	Rm 714 Main	800	" "

EVALUATION FACTORS	YES or NO	KEY	✓						
1. 5% or more of users* total man-hours (minimum 1 man-year) are being spent in looking up, searching, extracting, or correlating information or data. (*Users include all persons who personally do the looking up, searching, extracting or correlation. EXCEPT those assigned to operate the Information Facilities)	Y	-	Y	Y	Y	Y	N	N	N
2. Current information facilities are INADEQUATE for one or more of the following reasons: (*Circle any that apply*) A. Pertinent documents or information are regularly being missed or system produces too much non-relevant material or information. B. System can furnish documents, only, whereas users would like to receive only portions thereof, or precise data. C. System cannot satisfy need for retrieving precise data and correlating it.	Y	-	Y	Y	N	N	Y	Y	N
3. Much faster retrieval speed is needed than could ever be achieved under present or any other conventional method.	N	Y	N	N	N	N	N	N	N
4. Time presently spent in looking up, searching, extracting, or correlating information or data is mainly attributable to limitations of conventional methods.	Y	-	Y	N	Y	N	Y	N	-
CONCLUSIONS									
A. Modern information retrieval system seems a likely possibility		X	X		X				
B. Likely that present or improved conventional methods will suffice.						X		X	X
C. Likely that present or improved conventional methods will suffice; HOWEVER, also consider modern information retrieval systems (*Particularly those which use inexpensive tools*)				X			X		
D. Other (*Specify and explain*)									

INCONVENIENT FEATURES (*Features NOT necessarily attributable to limitations of conventional methods. CHECK ANY THAT APPLY.*)		
	DIFFICULT TO OBTAIN ACCESS TO INFORMATION	
	USERS PREFER TO SEARCH BUT FIND SYSTEM DIFFICULT TO UNDERSTAND OR USE	
	USERS NOT ROUTINELY INFORMED OF NEW INFORMATION PERTAINING TO THEIR WORK	✓
	OTHER (*Specify and explain*) Records too far removed from users	

REMARKS

Unknown factors are the possible losses in operating efficiency, program effectiveness and manpower management caused by the inadequacies & limitions of present system.

INSTRUCTIONS - Prepare as many of these Decision Tables as needed to collect data during the course of surveying individual user groups to estimate manhours spent in looking up, searching, extracting, or correlating information or data. Summarize your findings by preparing one Decision Table for each of the broad, similar types of information required at the installation being surveyed.
Enter "YES" or "NO" in the column opposite each of the Evaluation Factors to indicate existing conditions.
Compare your overall findings with those in the columns under "KEY" until you find a set that matches yours - place a check mark at the top of that column (*preferably with a colored pencil*). Follow the selected column down to the "CONCLUSIONS" and circle the appropriate X.

Figure 37

- To serve as a ready reminder of future action that should be taken in addition to or independent of the installation of a retrieval system.

- To supplement the data in evaluation factors 2 and 4 in borderline situations by providing additional clues as to which system to select—a conventional or a modern information retrieval system.

All of the inconvenient features listed could probably be corrected by adjusting and improving the existing conventional system.

Summary

The forms shown in this chapter, like all the others appearing in this handbook, are offered as suggested working tools only, to be used by those conducting the information retrieval studies. They are designed to assist in data gathering, analysis, decisionmaking, and documentation of the study. The forms may be used in their present format or may be modified to suit the needs of individual agencies.

The decision tables are not intended to substitute for human judgment, but rather to aid in quickly identifying those situations where a modern information retrieval system may be justified. In order to apply them correctly, it is not only necessary to fully understand how they are to be used, as explained in this chapter, but also to have a comprehensive knowledge of the limitations and advantages of conventional systems. This was discussed briefly in chapter I; if, however, the person conducting the study has not had experience in designing and operating conventional filing and library systems, additional research in these areas should be conducted. It is recommended that the National Archives and Records Service (NARS) records management handbooks *Subject Filing*, *Files Operations*, and *File Stations* be reviewed, in any event, before undertaking the preliminary survey.

When conducting a preliminary survey, the study should begin with a look at the information facilities. However, the findings should be organized on the basis of the broad types of information needed rather than by organizational elements or file stations. The reason for this is that only in rare instances is any particular type of information of interest to only a single organizational element. Further, the information is often drawn from more than one source, and the same information is usually found in more than one information facility.

The person conducting the survey should identify the broad types of information needed by the users as early as possible and then relate to each type the user groups and the file stations that serve as the source of the information. The final decision as to whether there is a potential need for an information retrieval system thus takes into consideration the varying needs of individual user groups as well as problems incurred in the operation of the information facility.

The data gathered and the conclusions reached during the preliminary survey are not of course adequate for going ahead and installing a system. A large scale information retrieval study and system installation might typically consist of the following phases:

1. The preliminary survey
2. Determination of system requirements (the feasibility study)
3. Development of system concepts and preliminary system design
4. Determination of equipment requirements and selection of equipment
5. Development of detailed system design and recruitment of personnel
6. Acquisition of equipment and training of personnel
7. Implementation and testing of equipment and orientation of users
8. Evaluation of system performance, and periodic revision of system

This handbook does not attempt to cover all these phases, but instead concentrates on those matters peculiar to information retrieval or those presenting special problems in designing, installing, and operating an information retrieval system.

VII. HOW TO DETERMINE SYSTEM REQUIREMENTS

The data gathered during the preliminary survey is far too sketchy and unreliable to serve as the basis for determining system requirements. Consequently, it is necessary to go back to those areas where there was an apparent potential need for modern information retrieval methods and to obtain additional data in order to make a further, more detailed analysis.

Data Collection Techniques

The various techniques that might be used in collecting the data are described below. These techniques are intended to complement rather than duplicate each other, although some redundancy is always desirable in order to verify the findings. In a large scale study, all or most of these techniques might be employed. However, there will always be situations where the use of a certain technique is not permissible or perhaps not practical or necessary. The objective of the person conducting the study should be to obtain the needed data in the best way possible to assure its completeness and accuracy and at the same time to minimize interruptions in the work of the organization and the man-hours expended by users and others involved in the study.

Questionnaires. Questionnaires, although not an entirely reliable or satisfactory method for gathering data, can be quite helpful, particularly in the area of user needs. Considerable care and testing are needed in phrasing the questions and interpreting the results in order to avoid misleading or invalid conclusions.

Interviews. Some of the information will necessarily be obtained through interviews. Interviews are also a good way to gain an understanding of the working climate and the attitudes of the individuals and to follow up on questionnaires when necessary.

Observations. Some of the data needed to determine system requirements can be ol through on-site observations. Data such a rent file size, physical characteristics of tl ords, and the age of the current collection i obtained in this manner. Personal observa needed to ensure a good understanding of t uation and can also serve as a check again obtained through questionnaires and interv

Reports. The questionnaires, interview observations will not provide all the data n Data such as work volume, man-hours us record inventories may appear in existing r Consequently, the person conducting the should look over the existing reports and them whenever possible for obtaining data. Also, of course, data gathered in con with the preliminary survey should be used phase of the study.

Work Counts. While work counts sho used sparingly, they may be essential for c ing data not contained in any existing repo available through other sources. The work may be needed to obtain or verify such dat put volume, man-hour requirements, tin number of searches, average searching tim volume of information retrieved. The pe the work count will vary according to the p lar situation, but normally it should not r be longer than 30 days; such counts shou ploy sampling techniques rather than atte to be a 100 percent check. In a large-scale consideration should be given to the use of anized techniques employing source data a tion (SDA).

Suggested Questionnaires

Figures 38 and 39 are examples of questio that might be employed for collecting in tion regarding user needs. Both the items form and question sections would mor likely have to be modified or rephrased t the questionnaires to the particular organ under study.

User's Report, Information Requirements, General (Figure 38). A questionnaire such as this one might be used to obtain an overall picture of the user needs, work habits, preferences, information problems, and recommendations. Consequently, it tends to be complex and would probably require somewhat detailed explanations and examples of answers appropriate under various circumstances. A brief orientation, preferably through group discussion, is therefore needed in order for the users to properly understand the questionnaires and thus obtain worthwhile results. This orientation should be part of the "Users' Briefing" described later in this chapter—another good reason such a briefing is highly desirable.

A review of this questionnaire reveals that it is used to probe for facts that will have a vital impact on the design of any information system. The answer to question 8 may of necessity be only an estimate, unless there is sufficient time and need for requesting selected users to maintain a diary (daily log) for a specified period. Some of the questions are purposely redundant to a certain extent in that essentially the same information is occasionally asked for in different ways since some of the questions will not be fully understood by all the users.

It should be expected that the cooperation and quality in completing the questionnaires will range from very good to very poor—therefore, those conducting the study must be careful not to jump to conclusions but instead should give careful thought to the circumstances, environment, biases, and other factors that may have affected the way the questionnaires were completed.

Follow-up interviews are absolutely essential to effective use of the questionnaires. Interviews should be conducted for clarification of significant inconsistencies or errors and when a user obviously needs assistance in completing a questionnaire. Some questions, such as numbers 4, 7, and 9-16, may be designed to produce clues rather than complete answers and explanations; and therefore, these queries require follow-up discussions with individual users to obtain a full understanding of the situation and its possible impact on an information retrieval system.

User's Report, Work Unit Information Requirements (Figure 39). This second questionnaire might be used to obtain an across-the-board sampling of actual current information needs and user practices. It is designed to find out how the user goes about getting the information needed to complete a specific task, for example, processing a case, answering an inquiry, making a study, or writing a new procedure.

To decide how many tasks or work units are to be reported the following guidelines are suggested:

1. If the nature of the work is such that more than one task or work unit is completed each day, request the users to prepare five forms, i.e., one for the first task performed each day for the next five days after the briefing session.

2. If the individual task or work unit varies in length from one to five days, have the user report only on the first new task occurring after the briefing session.

3. If the individual tasks or work units are usually longer than five days, complete the form to show the related information activities for a one-week period or upon completion of the task, whichever occurs first.

It is also necessary to determine whether the questionnaire will be distributed to each user or only to certain ones. Whenever possible, most of the users should be asked to complete them. The three categories of information in this questionnaire are:

- Questions 1–5 seek information about the nature of the task, the end product, the character of the information needed, the way in which the user identified it, and where he went to get the information.

- Questions 6–8 cover information on how the user went about getting the needed information, the techniques used, and the man-hours involved.

SAMPLE QUESTIONNAIRE FOR USER'S REPORT ON GENERAL INFORMATION REQUIREMENTS

USER'S REPORT INFORMATION REQUIREMENTS, GENERAL	Complete each of the questions to the best of your knowledge. Enter N/A for questions, when not applicable.	
1. NAME	JOB TITLE	
2. PRIMARY DUTIES OR RESPONSIBILITIES		3. LENGTH OF TIME IN THIS WORK

4. ANY SEASONAL OR OTHER PERIODIC PEAK PERIODS, WHEN INFORMATION NEEDS TEND TO INCREASE?

☐ NO ☐ YES *(identify)*

5. ARE THE TASKS AND INFORMATION OR DATA REQUIREMENTS DESCRIBED IN THE ATTACHED "USER'S REPORT, WORK UNIT INFORMATION REQUIREMENTS" TYPICAL?

☐ NO ☐ YES *(explain)*

6. HOW IS THE INFORMATION OR DATA GENERALLY USED IN COMPLETING YOUR WORK ASSIGNMENT?

☐ DIRECTLY INCORPORATED IN THE END PRODUCT ☐ OTHER

☐ DIRECTLY INCORPORATED IN THE END PRODUCT AS BACKGROUND INFORMATION

7. DO YOU PREFER TO DO YOUR OWN SEARCHING OR INFORMATION LOOK-UP, RATHER THAN HAVING SOMEONE OR A MACHINE DO IT FOR YOU?

☐ NO ☐ YES *(explain)*

8. CHECK APPROPRIATE BOXES AND COMPLETE ITEMS BELOW TO INDICATE NET TIME PERSONALLY SPENT IN OBTAINING INFORMATION AT AN INFORMATION FACILITY *(include time spent at your desk or work station and personal files)*

CHECK	ITEM	TITLE OF INFORMATION FACILITY OR SOURCE	LOCATION	MONTHLY ACTIVITY	
				NO. OF TIMES	TOTAL HOURS
	GENERAL SEARCHING FOR INFORMATION CONTAINED IN WRITTEN TEXT				
	RETRIEVAL OF SINGLE SENTENCES, PARAGRAPHS OR OTHER STATEMENTS CONTAINED IN WRITTEN TEXT				
	RETRIEVAL OF GRAPHIC OR PICTORIAL MATTER				
	LOOKING UP, COPYING, EXTRACTING, OR FURNISHING DISCRETE DATA *(such as names, numbers, dates, and quantitative or qualitative data)*				
	LOOKING UP, CORRELATING, COMPARING, REARRANGING OR OTHERWISE MANIPULATING DISCRETE DATA				
	SCANNING PERIODICALS, REPORTS, AND OTHER MATERIAL TO KEEP ABREAST OF LATEST DEVELOPMENTS IN YOUR FIELD				

9. GENERALLY ARE PRESENT RETRIEVAL SPEEDS ADEQUATE FOR NEEDS? ☐ NO ☐ YES	TIME FACTORS:	REQUIRED	DESIRABLE

IF ANSWER ABOVE IS "NO", IDENTIFY INFORMATION THAT NEEDS RETRIEVAL TIME REDUCED

Figure 38

SAMPLE QUESTIONNAIRE FOR USER'S REPORT ON GENERAL INFORMATION REQUIREMENTS

10. ARE YOUR RETRIEVAL EFFORTS HAMPERED BY ANY OF THE FOLLOWING CONDITIONS? *(check appropriate item boxes)*

- ☐ FILING AND/OR INDEXING NOT KEPT CURRENT
- ☐ SUBJECT CLASSIFICATION OR INDEXING SYSTEM IS INEFFECTIVE
- ☐ INSUFFICIENT INFORMATION OR DATA COVERAGE
- ☐ MORE THAN A FEW DOCUMENTS CONTAINING RELEVANT INFORMATION OR DATA ARE BEING MISSED
- ☐ DIFFICULT TO GAIN PHYSICAL ACCESS TO INFORMATION OR DATA
- ☐ A GREAT DEAL OF THE INFORMATION OR DATA FOUND IS USELESS OR REDUNDANT
- ☐ PHYSICAL FORM OR FORMAT OF MATERIAL IS INCONVENIENT
- ☐ ARRANGEMENT OR FILE SEQUENCE OF NON SUBJECT TYPE FILE IS NOT WELL SUITED TO YOUR NEEDS
- ☐ FILE NOT READILY BROWSABLE
- ☐ OTHER PROBLEMS
- ☐ SUBJECT CLASSIFICATION OR INDEXING SYSTEM DIFFICULT TO UNDERSTAND OR USE

11. WHAT HAS BEEN THE EFFECT OF THE ABOVE PROBLEMS ON YOUR WORK AND THE EFFICIENCY OF THE OFFICE?

12. WHICH OF THE FOLLOWING CONDITIONS MOST CLOSELY CORRESPONDS TO THE SEARCH RESULTS YOU NEED WHEN RETRIEVING INFORMATION BY SUBJECT?

- ☐ RETRIEVAL OF ALL DOCUMENTS OR OTHER RECORDS THAT MIGHT BE CONSIDERED RELEVANT TO THE QUERY WITH THE POSSIBILITY THAT A CONSIDERABLE AMOUNT MAY PROVE TO BE NONRELEVANT
  (to avoid the possiblity of overlooking any relevant material)
- ☐ RETRIEVAL OF ONLY THOSE DOCUMENTS CONTAINING THE SPECIFIC INFORMATION OR DATA DESCRIBED IN THE QUERY WITH THE POSSIBILITY THAT DOCUMENTS OF VARYING DEGREES OF RELEVANCE MAY HAVE BEEN MISSED
  (to avoid retrieving more material than is really needed or can be readily used.)
- ☐ OTHER *(explain)*

13. WHICH ONE OF THE FOLLOWING DEGREES OF SPECIFICITY OR DEPTH OF SUBJECT MATTER BREAKDOWN FOR WRITTEN INFORMATION MOST CLOSELY CORRESPONDS TO YOUR NEEDS?

- ☐ LOW SPECIFICITY - BROADER THAN THE SUBJECT BREAKDOWN IN THE TABLE OF CONTENTS OF A TEXT BOOK OR MANUAL
- ☐ MODERATE SPECIFICITY - ROUGHLY EQUIVALENT TO THE SUBJECT BREAKDOWN IN THE TABLE OF CONTENTS OF A TEXT BOOK OR MANUAL
- ☐ HIGH SPECIFICITY - MORE SPECIFIC THAN THE TABLE OF CONTENTS OF A TEXT BOOK OR MANUAL

14. ARE THERE ANY PARTICULAR FUNCTIONS OR WORK PERFORMED BY YOUR ORGANIZATION, WHICH YOU BELIEVE COULD BE SUBSTANTIALLY IMPROVED OR PERFORMED AT LESS COST THROUGH THE APPLICATION OF MODERN INFORMATION RETRIEVAL TECHNIQUES?

☐ NO ☐ YES *(explain)*

15. ARE THERE ANY PARTICULAR TYPES OF ARTICLES, REPORTS, OR OTHER RECURRING MATERIAL ABOUT WHICH YOU NEED TO BE ROUTINELY INFORMED ABOUT IN ORDER TO KEEP ABREAST OF THE LATEST DEVELOPMENTS IN YOUR FIELD?

☐ NO ☐ YES *(describe)*

16. ARE THERE ANY INFORMATION OR DATA FILES NOW BEING MAINTAINED, WHICH MAY NOT BE NEEDED, IF AN INFORMATION RETRIEVAL SYSTEM WERE INSTALLED?

☐ NO ☐ YES *(identify)*

Figure 38 (Continued)

SAMPLE QUESTIONNAIRE FOR USER'S REPORT ON WORK UNIT INFORMATION REQUIREMENTS

USER'S REPORT

WORK UNIT INFORMATION REQUIREMENTS

Complete each of the questions to the best of your knowledge. Enter N/A for questions, when not applicable.

NAME	JOB TITLE	DATE

1 DESCRIBE THE TASK OR WORK UNIT THAT CREATED THIS NEED FOR INFORMATION

2 WHAT WAS THE OUTPUT?

3. PRIMARY CHARACTER OF INFORMATION SOUGHT:

☐ WRITTEN INFORMATION *(correspondence, directives, reports, and publications)*

☐ GRAPHIC OR PICTORIAL MATTER *(maps, drawings, and photographs)*

☐ QUANTATIVE QUALITATIVE AND OTHER DATA *(statistics, performance standards, costs, size, color, shape, etc.)*

☐ OTHER *(specify)*

4 WHAT PARTICULAR IDENTIFYING FEATURE, DESCRIPTIVE TERM, OR OTHER MEANS SERVED AS THE PRIMARY BASIS FOR IDENTIFYING THE INFORMATION OR DATA SOUGHT *(specify whether primarily documents or record title; name or numbers, names, titles or numbers or other people, places or things, subject topics; quantitative data, etc.)*

5 IDENTIFY THE INFORMATION OR DATA FACILITIES OR SOURCES USED INCLUDING PERSONAL FILES

TITLE OF INFORMATION SOURCE	ORGANIZATIONAL AND PHYSICAL LOCATION	PHYSICAL FORM OF MATERIAL

6 WAS ASSISTANCE RECEIVED? ☐ NO ☐ YES

IF YES, ENTER NAMES OR TITLE AND ORGANIZATIONAL LOCATION

7 TYPE OF PERSONAL SEARCHING PERFORMED	MANHOURS SPENT
GENERAL SEARCHING FOR INFORMATION CONTAINED IN WRITTEN TEXT	
LOOKING UP COPYING EXTRACTING OR FURNISHING DISCRETE DATA	
LOOKING UP CORRELATING COMPARING REARRANGING OR OTHERWISE MANIPULATING DISCRETE DATA	
OTHER *(specify)*	

8 METHODS AND MATERIALS USED IN PERSONAL SEARCHING:

☐ USED SUBJECT INDEX CARDS

☐ USED PRINTED INDEX AND/OR TABLE OF CONTENTS

☐ BROWSED ENTIRE DOCUMENT FILE

☐ SCANNED CONTENTS OF FOLDERS OR OTHER DOCUMENTS ARRANGED BY SUBJECT TOPICS

☐ SCANNED CONTENTS OF FOLDERS OR OTHER DOCUMENTS ARRANGED BY NAMES OR NUMBERS OF PEOPLE, PLACES, OR THINGS

☐ OTHER *(specify)*

9 HOW LONG WAS IT FROM THE TIME YOU INITIATED ACTION TO GET THIS INFORMATION OR DATA UNTIL YOU OBTAINED IT?

10 HOW QUICKLY DID YOU ACTUALLY NEED THIS INFORMATION INFORMATION OR DATA?

11 TOTAL MANHOURS YOU PERSONALLY SPENT ON COMPLETING THE TASK OR WORK UNIT INCLUDING THE TIME SPENT ON OBTAINING INFORMATION OF DATA?	HOURS	MINUTES

12 HOW SUCCESSFUL WERE YOU IN OBTAINING THE NEEDED INFORMATION OR DATA?

☐ OBTAINED ALL OR MOST OF IT

☐ IMPORTANT INFORMATION OR DATA APPEARS TO HAVE BEEN MISSED

☐ INFORMATION OR DATA WAS NEVER FOUND

☐ OTHER *(specify)*

13 NEED FOR THIS INFORMATION:

☐ CRITICAL

☐ SIGNIFICANT

☐ MARGINAL

☐ OTHER *(specify)*

Figure 39

- Questions 9–13 request information about the quality of the search results and the relative importance of the information search to the overall completion of the task.

As in the case of the earlier questionnaire, there will be instances where it may be necessary or desirable to interview individual users to obtain additional information.

Data Summarization Techniques

As explained in chapter VI, the study findings should be organized on the basis of the types of information needed and then related to the user groups and the information facilities that serve as the source of the information. A form similar to the system requirement worksheet (figure 40) may be used for this purpose. Such a form can serve not only as a convenient means for organizing the data but also as a checklist to assure that nothing of significance has been overlooked. One system requirement worksheet should be prepared for each of the broad types of information needed by the installation under study.

The sample system requirement worksheet is divided into four parts, as follows:

Part A—Input and Storage, page 1.

Part B—Retrieval and Presentation, page 2.

Part C—Resources, pages 3 and 4.

Part D—General Improvements Needed, page 4.

In conducting the study, of course, the output requirements for the system must be determined before it can be decided what information will have to be stored. Consequently the data for part B, retrieval and presentation, would have to be gathered first or perhaps simultaneously with that for part A, input and storage. While the form is largely self-explanatory, the following notes are offered to assist in its use.

Part A, Input and Storage. In examining input and storage requirements, the nature and volume of material that would have to be entered into must be known; therefore, this part reflects not only the current situation but future expectations as well.

Item 1, Physical characteristics. The physical characteristics of the input must be known since they have a direct effect on the type of equipment that can be used and personnel requirements.

Item 2, File size factors. Since some methods and equipment have optimum limits on the volume of material that can be stored or involve high storage costs, file size is always an important factor.

Item 3, Intellectual characteristics. Knowledge of the intellectual characteristics is needed since the more complex the intellectual requirements, the more sophisticated the system may have to be.

Item 4, Source factors. The source factors, like physical characteristics, directly affect ease of input and the type of storage equipment. For example, if the documents or data are produced in-house and could be received in computer magnetic tape form, the possibilities would be quite different from those where the producer is an outside organization and the information is available in printed form only.

Item 5, Change factors. If changes to the information entered into the system will be necessary, this fact must be known, since making the changes could be difficult and expensive if certain methods and equipment were to be employed.

Part B, Retrieval and Presentation. In this part are compiled the data needed to provide a comprehensive summary of user needs.

Item 1, Search activity factors. Types of retrieval actions and volume are important factors, since there are usually practical limitations in the workload that each equipment class can handle. The location of the users and their proximity to each other are also factors that might cause one method or type of equipment to be impractical and another to be ideally suited to the situation at hand.

Item 2, Search intellectual characteristics. If the users ask for documents or data by case name or number, the intellectual requirements imposed on

SAMPLE FORM FOR SYSTEM REQUIREMENTS WORKSHEET

SYSTEM REQUIREMENT WORKSHEET

TYPE OF INFORMATION	NAMES AND LOCATIONS OF INFORMATION FACILITIES INVOLVED

PRIMARY TYPE OF DOCUMENT OR RECORD:

☐ WRITTEN INFORMATION ☐ GRAPHIC OR PICTORIAL ☐ PRECISE DATA (maps, numbers, dates, etc.) ☐ OTHER (*specify*)

PART A - INPUT AND STORAGE

1. PHYSICAL CHARACTERISTICS (*Document or record to be entered into the system*)

a. DOCUMENT SIZE AND FORM (*3x5" cards, 8x10½" sheets, 6x9" bound books, etc.*)	
b. AVERAGE LENGTH OF INPUT DOCUMENT OR RECORD SEGMENT (*number of page, if text; number of characters or lines, if data, etc*)	
c. LENGTH OF LONGEST INPUT DOCUMENT OR RECORD SEGMENT (*number of pages, if text, number of characters or lines, if data, etc*)	

2. FILE SIZE FACTORS (*of information or data to be stored*)

	NO. OF DOCUMENTS OR RECORD UNITS	TOTAL PAGES, CHARACTERS OR LINES
a. PRESENT QUANTITY		
b. ESTIMATED QUANTITY IN TWO YEARS		
c. ESTIMATED QUANTITY IN FIVE YEARS		
d. MONTHLY GROWTH		

e. OBSOLESCENSE FACTOR (*period or event after which document or record is no longer needed*)

f. AGE OF CURRENT COLLECTION

3. INTELLECTUAL CHARACTERISTICS (*of information or data to be stored*)

a. ESTIMATED TOTAL NUMBER OF INDEXING TERMS, ATTRIBUTES, ETC. IN INDEX VOCABULARY	IF POSSIBLE, ESTIMATE AVERAGE NO. ASIGNED EACH DOCUMENT OR RECORD

b. STABILITY OF FIELD: ☐ STABLE ☐ SMALL AMOUNT OF CHANGE ☐ CONSTANTLY CHANGING

c. DEGREE OF COMPLEXITY: ☐ SIMPLE DATA ☐ COMPLEX DATA ☐ ORDINARY TEXT ☐ COMPLEX TEXT

d. SCOPE AND RANGE OF COVERAGE: ☐ NARROW ☐ MEDIUM ☐ BROAD

e. MEANS USED TO IDENTIFY AND OR DESCRIBE DOCUMENT OR RECORDS PRIOR TO RECEIPT BY FACILITY (*title or number, author, abstract, case or record number, etc.*)

f. DEGREE OF REDUNDANCY OF INFORMATION OR DATA WITHIN THE FILE: ☐ LOW ☐ MODERATE ☐ HIGH

g. EXTENT OF INPUT EVALUATION OR SCREENING NEEDED

4. SOURCE FACTORS (*of information or data to be stored*)

a. ABILITY OF SOURCE TO FURNISH IDENTIFYING, DESCRIPTIVE, OR OTHER DATA IN MACHINEABLE FORM

b. EXTENT OF DUPLICATION

DUPLICATED DATA	HOW MUCH	LOCATION	TYPE OF SYSTEM	COMPATIBILITY WITH SYSTEM UNDER CONSIDERATION

c. PRIMARY PRODUCERS OF DOCUMENTS OR DATA

IDENTITY	LOCATION

5. CHANGE FACTORS

a. EXTENT TO WHICH INDIVIDUAL DOCUMENTS, THEIR DESCRIPTIONS, OR IDENTIFYING TERMS, AND OR DATA TO BE STORED IN THE SYSTEM WILL HAVE TO BE CHANGED, UPDATED, ADDED TO, OR DELETED

Figure 47

SAMPLE FORM FOR SYSTEM REQUIREMENTS WORKSHEET

PART B - RETRIEVAL AND PRESENTATION

1. SEARCH ACTIVITY FACTORS

a. VOLUME — MONTHLY RETRIEVAL ACTIVITY

NO. OF USERS	ORGANIZATION AND LOCATION	TYPE OF ACTIONS	NO.	NO. OF DOCUMENTS, PAGES OR CHARACTERS RETRIEVED

b. PHYSICAL DISPERSION OF USERS (*in percentages*)	SAME FLOOR, SAME BUILDING %	SAME BUILDING %	SAME BLDG. COMPLEX %

c. SEARCH ACTIVITY PERIODIC FLUCTUATIONS - IF ANY (*describe*)

2. SEARCH INTELLECTUAL CHARACTERISTICS

a. WRITTEN INFORMATION

SEARCH SPECIFICITY:	EXTENT OF CORRELATION REQUIRED (*subject*)	AVERAGE NO. OF INDEX TERMS USED PER SEARCH
☐ LOW ☐ MODERATE ☐ HIGH	☐ NONE ☐ LOW ☐ MODERATE ☐ HIGH	

b. DATA RETRIEVAL

COMPLEXITY:	EXTENT OF CORRELATION REQUIRED (*data*)	AVERAGE NO. OF DATA ITEMS PER RETRIEVAL ACTION
☐ SIMPLE ☐ MODERATELY COMPLEX ☐ COMPLEX	☐ NONE ☐ LOW ☐ MODERATE ☐ HIGH	

c. IDENTIFYING FEATURES TO BE USED TO IDENTIFY INFORMATION (*name, number, document or record title, place or thing, attribute or other index term, etc.*)

d. OTHER (*specify*)

3. OUTPUT OR PRESENTATION, (*physical characteristics needed OR DESIRED*)

a. TYPE OF OUTPUT (*selected data or facts, document no's., whole documents, selected portions, etc.*)

b. METHOD OF PRESENTATION OR DISPLAY (*manual display of document no's. or index records, etc.*)

c. OTHER (*describe*)

4. SERVICE REQUIREMENTS

a. SPEED

MAXIMUM PERMISSIBLE PERIOD BETWEEN TIME INFORMATION, DATA, OR RECORD FIRST REQUESTED OR NEEDED AND TIME DELIVERED	MAXIMUM BATCHED PROCESSING OF REQUESTS PERMISSIBLE (*specify daily or weekly*)	IF A SEPARATE DOCUMENT REFERENCE INDEX FILE WERE TO BE USED, GIVE MAXIMUM PERMISSIBLE TIME FOR: CONDUCTING A SEARCH	DELIVERY OF A DOCUMENT

b. CONVERTABILITY AND COMPATABILITY WITH OTHER AGENCY SYSTEMS & EQUIPMENT (*if essential - describe*)

c. ALTERNATE SEARCH METHODS (*if needed - explain*)

d. USER SELF SEARCHING: ☐ NONE ☐ DESIRABLE ☐ ESSENTIAL

e. BROWSABILITY (*describe special needs, if any*)

f. CURRENT AWARENESS (*if needed or desired, indicate type and frequency of service*)

g. OTHER SERVICE REQUIREMENTS (*describe*)

5. QUALITY REQUIREMENTS

a. ACCURACY OF EQUIPMENT OR DEVICE (*explain*)

b. EQUIPMENT RELIABILITY ☐ NORMAL ☐ CRITICAL

c. PROTECTION AGAINST LOSS OF INFORMATION OR DATA STORED:
☐ ROUTINE PRECAUTIONS
☐ SPECIAL MEASURES NEEDED

d. WRITTEN INFORMATION RECALL - PRECISION RATIO:
☐ HIGH RECALL ☐ OTHER (*specify*)
☐ HIGH PRECISION

e. CURRENCY FACTOR (*specify how up to date the information must be*)

f. OTHER QUALITY REQUIREMENTS - IF ANY

Figure 40 (*Continued*)

SAMPLE FORM FOR SYSTEM REQUIREMENTS WORKSHEET

PART C - RESOURCES

CURRENT ANNUAL INFORMATION COST (Complete this part to summarize total current information retrieval resources and costs, which need to be taken into consideration, when developing the proposed system. Enter N/A for any item not applicable.

1. PERSONNEL COSTS	a. SUPERVISION AND OPERATION OF INFORMATION FACILITIES	
	b. OTHER PERSONNEL COST INVOLVED IN STORING AND RETRIEVING THIS DATA	
	TOTAL ANNUAL PERSONNEL COSTS	
2. EQUIPMENT COSTS AND SERVICE CHARGES	a. RENTAL COSTS, IF ANY	
	b. DEPRECIATION	
	c. MAINTENANCE	
	d. OTHER EQUIPMENT COSTS	
	TOTAL ANNUAL EQUIPMENT COSTS	
3. SUPPLY COSTS	a.	
	b.	
	c.	
	TOTAL ANNUAL SUPPLY COSTS	
4. SPACE AND MISCELLANEOUS COSTS	a. SPACE	
	b. MISCELLANEOUS COSTS	
	TOTAL ANNUAL SPACE AND MISC. COSTS	
5.a TOTAL GROSS CURRENT ANNUAL COSTS to be taken into consideration in development of the proposed system(*items 1 through 4*)		
5.b LESS ANNUAL RESIDUAL COSTS, IF ANY FOR SERVICING EXISTING INFORMATION FACILITIES, WHICH MUST STILL BE REFERRED TO AFTER NEW SYSTEM IS INSTALLED		
6. ADJUSTED GROSS CURRENT ANNUAL EXPENDITURES FOR STORAGE AND RETRIEVAL OF INFORMATION, WHICH ARE AVAILABLE FOR APPLICATION TOWARD COST OF NEW SYSTEM		
7. ESTIMATED COSTS attributed to NOT being able to RETRIEVE and/or manipulate information WHEN NEEDED. *(describe)*		
8. VALUE OF USER MANHOURS, which could be saved, if modern information retrieval system was installed		
9. TOTAL Estimated net annual expenditures, which are available for application to cost of proposed information retrieval system (*add items 6 through 8*)		

CURRENT CAPABILITY

1. PERSONNEL

a. AVAILABILITY OF PERSONNEL TO DESIGN, INSTALL, AND PROVIDE TECHNICAL SUPERVISION OF AN INFORMATION RETRIEVAL SYSTEM

PRESENT JOB TITLES	GRADE	EDUCATION, TRAINING, AND EXPERIENCE	PRESENT ASSIGNMENT

Figure 40 (Continued)

SAMPLE FORM FOR SYSTEM REQUIREMENTS WORKSHEET

1. PERSONNEL *(Continued)*

b. AVAILABILITY OF PERSONNEL TO OPERATE AN INFORMATION RETRIEVAL SYSTEM

PRESENT JOB TITLES	GRADE	EDUCATION, TRNG. AND EXP.	PRESENT ASSIGNMENT

2. MECHANIZED EQUIPMENT

a. AVAILABLE FULL TIME

QUANTITY	NAME AND MODEL	LOCATION	OWNED OR RENTED

b. AVAILABLE PART TIME *(use the same column heads above)*

QUANTITY	NAME AND MODEL	LOCATION	PERIOD HOURS

3. SPECIAL INFORMATION AIDS OR TOOLS

DESCRIBE ANY SPECIAL AIDS OR TOOLS CURRENTLY USED IN STORING AND RETRIEVING INFORMATION, OR DATA, FOR EXAMPLE SUBJECT CLASSIFICATION OUTLINES; INDEX VOCABULARIES; DATA TABLES, ETC.

DESCRIPTION	OFFICE OF ORIGIN OR JURISDICTION

4. PHYSICAL AND OTHER FACILITIES

a. COMMUNICATION AND TRANSPORTATION *(mail, teletype, messenger service, conveyors, shuttle bus, etc.)*

DESCRIPTION	SPEED FREQUENCY	COST

b. SPACE

QUANTITY AND LOCATION	AIR CONDITIONED	GOOD LIGHTING

c. ELECTRICAL POWER

CAPACITY	RESTRICTIONS, IF ANY

PART D GENERAL IMPROVEMENTS NEEDED (Prepare this form to indicate improvements, which should or could be made, regardless of whether or not a modern information retrieval system is to be installed. Identify the information facilities and groups involved in each instance.)

1 REDUCTION IN TIME LAG:
- ☐ INPUT - REDUCE DELAYS IN ENTERING INFORMATION
- ☐ OUTPUT - REDUCE DELAYS IN MAKING SEARCH OR DELIVERY OF ITEMS TO THE USERS

2 STAFFING:
- ☐ OBTAIN SPECIALISTS
- ☐ CONDUCT ADDITIONAL TRNG.
- ☐ ASSIGN ADDITIONAL PERSONNEL OR MANHOURS

3 ORGANIZATION AND CONTROL OF INFORMATION:
- ☐ INCREASE INFORMATION OR DATA COVERAGE
- ☐ ELIMINATE USELESS OR REDUNDANT INFORMATION
- ☐ REARRANGE OR REORGANIZE CASE FILE
- ☐ UPDATE OR REVISE OBSOLETE SUBJECT FILE
- ☐ CONSOLIDATE INFORMATION FACILITIES

4 USE:
- ☐ IMPROVE PROCEDURES AND FORMS FOR OBTAINING DATA
- ☐ RELOCATE FACILITY FOR BETTER ACCESSIBILITY
- ☐ IMPROVE COMMUNICATIONS BETWEEN SYSTEM OPERATORS AND USERS
- ☐ USE ALTERNATE FACILITIES
- ☐ ORIENT USERS IN USE OF PRESENT FACILITIES

5 OTHER *(describe)*

Figure 40 (Continued)

the system will be practically nil. If, on the other hand, the users ask for the documents on the basis of the subject topics or attributes, the method and equipment must have quite another intellectual capability. It is usually wasteful and more expensive to acquire equipment that has "intellectual" ability far exceeding that which is actually needed; or, in the opposite situation, it would be a grave mistake to install a system that fails to fully satisfy complex needs.

Item 3, Output or presentation, physical characteristics needed or desired. If the users must have the entire document, the demands on the system and equipment would be quite different than in a situation where they want precise data or desire to have the answers presented in special printed form or on a cathode ray tube (CRT).

Item 4, Service requirements. It is the throughput speed, rather than the speed at which equipment internal processing takes place, that is important to the user. Also, it is important that the person making the study be aware of any need for making the system compatible with other systems and equipment that may presently exist or are planned for the future.

One must be aware of any alternate search methods that may be needed because some of the user groups are at remote locations or do not need a system having as much retrieval capability as other groups. It is important to know the extent of user self-searching as opposed to searching by an intermediary, since this will be of concern in selecting the right method and equipment.

The person designing the system must also know whether it must be "browsable"—i.e., permits the operator to scan or skim through the system freely and at the same time to see the results of his search, rather than having to formulate precise questions and to wait a considerable period for the answers. Further, it is necessary to know whether a need exists for incorporating a current awareness or selective dissemination of information (SDI) capability in the system to automatically notify or forward information to employees when it has a bearing on their area of interest. If such a capability must be included, this would also have an effect on the method and equipment to be used.

Item 5, Quality requirements. If the system is to be used for conducting subject searches, it must be known whether the system should have high recall; that is, retrieval of all information that might be in any way pertinent, or high precision; i.e., retrieval of only that information that has a high degree of pertinency. (See chapter IX). If the system should operate somewhere between the two, this too must be known when the system is designed.

Part C, Resources. The purpose of this data is to determine the extent to which the costs, equipment needs, and personnel requirements for a new information system could be offset by expenditures, equipment, and personnel now being expended for storage and retrieval of information.

Current annual information costs. The person conducting the study needs to ascertain which of the current personnel and other costs for operating present information facilities and conducting searches could be applied to offset the costs for a modern information retrieval system. This should also take into consideration savings of users' time made possible through the introduction of modern information retrieval methods.

Current capability. It is necessary to know whether there are people available who would be capable of designing, installing, and technically supervising a modern information retrieval system; for if such telent is not present or could not possibly be obtained, it would be senseless to recommend installation of the system. Similarly, the person conducting the study must also take into consideration the qualifications of the personnel and the capability of any equipment that would be available, particularly if the system will be used for subject-type retrieval.

Part D, General Improvements Needed. The purpose of gathering this data is to isolate and identify weaknesses or failures in the present system that are not necessarily the fault of the type of system in use, but rather the way it is being managed and operated. The person conducting the study should review these conditions carefully since they too would affect the design of a new system and present their own particular problems, some of which may be overlooked or ignored on the assumption that the new system will automatically solve them.

If conditions such as inadequate staffing, work backlogs, user resistance, and poor utilization of existing facilities persist under the present system, the same thing could occur if a modern information retrieval system were to be installed. It is imperative, therefore, to consider all future plans and proposals in the light of any needless weaknesses or failures in the past in order to gain the ability to prevent the same thing from happening if the new system were to be adopted.

Final Review and Analysis of Findings

After all the system requirement worksheets have been completed, a review should be made of the manner in which the information needs have been grouped. The scope and content of each of the broad types of information should be scrutinized for the purpose of determining whether any adjustments need to be made; for example, consolidation of two or more broad types into an even broader type.

This final analysis and review is very important, since each of these broad types of information represents, in effect, a separate "information center" and will be individually considered in initially selecting the methods and equipment to be used.

Users' Briefings

It is during the data gathering and analysis phase that the users should be brought into the picture. This has several advantages:

- First, gaining their interest and understanding helps assure better cooperation and thus achieves better results from the questionnaires.
- Second, the potential users, through a newly acquired knowledge of information retrieval, may come up with potential applications and ideas that would otherwise have escaped the attention of those conducting the study.
- Third, establishing an early working partnership with the users goes a long way toward reducing problems that are likely to occur in the installation stage—particularly those involving lack of user acceptance and understanding of the new system.

Consequently, one or a series of briefings should be conducted for those users who the preliminary survey indicates have a potential need for modern information retrieval methods. The briefing should consist of the following three parts:

1. Background information.
2. An introduction to modern information retrieval theory and methods.
3. Illustrated presentations or demonstrations of information retrieval methods and equipment.

Use of General Analysis Techniques and Tools

The special tools and guidelines featured in this handbook are intended to implement and not to replace those normally used in conducting systems studies. They are designed to assist in tailoring studies to the particular factors and considerations involved in information storage and retrieval. It may still be necessary, for example, to use spread sheets and matrixes to compile and display the data collected.

It may also probably be necessary to prepare process, flow, work distribution, or operation charts—in other words, to employ many of the same techniques and tools commonly used in conducting any methods and procedures or systems study, particularly those pertinent to ADP or mechanization feasibility and application studies.

VIII. SELECTING THE RIGHT METHODS AND EQUIPMENT

Because there are such a variety of methods and equipment used in information retrieval, selecting the right one is never a simple or easy task. The process starts with the elimination of those methods and equipment classes that are clearly not suitable or practical. It ends with the comparison of the system requirements for the job at hand against the capability, characteristics, costs, and other features of the remaining classes.

Step 1. Selecting the Applicable Functional Category

The first task in the selection process, elimination of those methods and equipment classes not suitable or practical, may be accomplished by determining exactly what information retrieval function or functions the proposed system must perform. Once this is done, the person conducting the study needs to be concerned only with those methods and equipment classes which are normally used to perform that function or functions. To make the task easier, this chapter identifies the various methods and equipment classes according to four broad functional categories as follows:

Document Reference (DR) Systems. These systems are used primarily for subject-type searches to identify documents, persons, places, or things that are pertinent to the search questions. The user or person conducting the search is given the name or number of the document, person, place, or thing; and he then refers to the complete document or record to find out the details. Such systems are intended to quickly reduce a mountain of information to a manageable handful.

One example of a system performing the DR function is an electronic computer used in legal research to identify by the case name earlier court cases involving the same points of law and a situation similar to the one at hand. Another example is an optical coincidence system that is used to quickly identify those employees in the organization who possess the necessary qualification, characteristics, or attributes for a vacant position or special assignment.

Document Storage (DS) Systems. These systems are concerned mainly with the physical means for storing documents; the documents are arranged by some simple means such as titles or numbers. These systems cannot be used for conducting subject-type searches, but instead require that the user have a prior knowledge of the name, identifying number, machine address, etc. used to identify the desired document.

An example of a system performing the DS function is the microfiche system used by the research and development community for storage and distribution of technical reports. Another example is a video tape system used for storing applications and other important papers relating to housing loans.

Unified Reference-Storage (URS) Systems. These systems are, in effect, a combination of the first two functional categories. These systems are used mainly in situations where there is an urgent need to view the pertinent documents at the same time a subject-type search is being conducted. An example of a system performing the URS function is a microfilm system with photo-optical code used for storing technical correspondence and conducting searches on the basis of subject topics, contract numbers, names of equipment manufacturers, addressees, correspondence symbols, etc.

Data Fact Retrieval (DFR) Systems. These systems instead of merely referring the user to the name or number of the person, place, or thing, give the user the precise data or facts he is seeking. DFR systems are of two types—simple data lookup and complex data retrieval.

An example of a system performing the simple data lookup DFR function might be a mechanized roll microfilm system storing servicemen's allotment data and employing an odometer-type device to aid the user in quickly locating data relating to an individual serviceman. An example of a complex DFR retrieval system would be a computer system that maintains a large amount of data about each employee and then is used to compare, manipulate, select, and print data when conducting searches and preparing reports.

The decision chart depicted in figure 41 is intended as an aid in selecting the right (applicable) functional category, particularly for those who are conducting an information retrieval study for the first time.

Step 2. Selecting the Right Methods and Equipment

The second step consists of matching the system requirements as reflected in the system requirement worksheet against method and equipment capability, characteristics, cost, and other factors, as shown in the Nonconventional Methods and Equipment Guide, Appendix "A." Both this and the decision chart, figure 41, are designed to serve as only guides for quickly narrowing the wide, diverse fields of nonconventional methods and equipment to those few types that would normally be best suited to meet a particular set of system requirements and help make a final selection.

The nonconventional methods and equipment guide is organized in the same manner as the system requirements worksheet:

Part A—Input and Storage

Part B—Retrieval and Presentation

Part C—Resources.

The headings at the top of the columns on the guide refer to classes of equipment (not of any particular manufacturer). Part C, resources, must by necessity be completed by the person conducting the survey and is therefore separate.

After determining the appropriate functional category as explained above, it should be necessary to consider only those classes of methods and equipment marked "X" or "-X" for that functional category in the block immediately below the class title of the method or equipment. (However, there may be exceptional circumstances when one of the undesignated classes of methods and equipment will apply.) An "X" in the functional category block signifies that the particular method and equipment class is generally well suited for performing that function. A "-X", on the other hand, indicates that the method and equipment class might possibly be used to perform that function, but there may be limitations or other reasons it is less than ideally suited to many situations. (Descriptions of the various methods and equipment classes are included in chapters III, IV, and V.)

Because there will rarely be a situation where there is a perfect match between system requirements and equipment capabilities and characteristics, there usually will be a number of "trade-offs" to analyze and weigh. In some instances, the nonconventional methods and equipment guide identifies capabilities in terms of "ideal," and the fact that the system requirements do not fall specifically within that range should not necessarily bar the use of that particular class, but instead may merely put it in the questionable category. Much of the success of any methods and equipment class, including those with a strong "yes," depends upon the ability of the system designer. A methods and equipment class that initially appears questionable may, through clever systems design, prove entirely satisfactory.

Some of the advantages of a particular equipment class will be offset or outweighed by its disadvantages, when its application to the situation at hand is considered. There may also be some situations where, due to an overriding need or other peculiarity, an equipment family other than that pinpointed by the tables may be more appropriate; however, the tables would still serve as a means for obtaining a summary of the capabilities, advantages, and disadvantages of other equipment.

After deciding which method and equipment would be best suited to meet the needs for each of the broad types of information needed by the users, the analyst should then examine the situation in terms of overall installation needs and

DECISION CHART FOR SELECTING THE RIGHT FUNCTIONAL CATEGORY

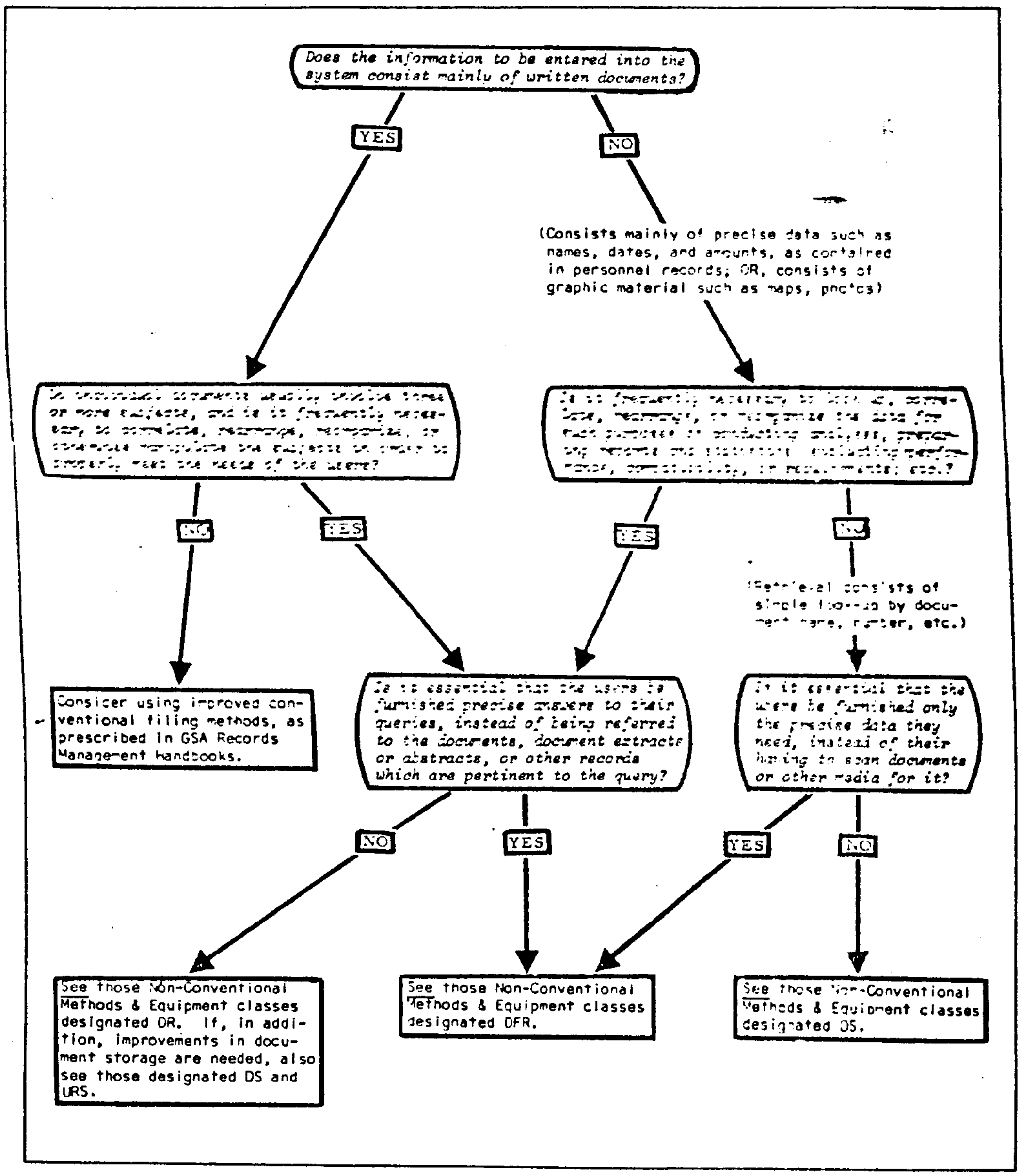


Figure 41

existing capability. The use of a reconciliation sheet similar to that shown in figure 42 is suggested for recording your findings and conclusions when matching individual system requirements against the capabilities, characteristics, and costs of the applicable methods and classes. It is suggested that the results be recorded as "yes" (Y), "no" (N) and "maybe" (?) in the blocks for parts A and B and also in the spaces for the overall conclusions. Part C, resources, of both the reconcilia-

SAMPLE RECONCILIATION SHEET FOR METHODS AND EQUIPMENT SELECTION

RECONCILIATION SHEET *SYSTEM REQUIREMENTS - METHODS AND EQUIPMENT*	TYPE OF INFORMATION (USER NEEDS)

INSTRUCTIONS - Follow instructions in Figure 1 and page E-1, and the check blocks at top of pages E-12 through E-15, of "Non Conventional Methods and Equipment Guide" to determine which method and equipment classes should be considered for storage and retrieval of the type of information (user needs) described above, and place an "X" [illegible] classes on this form. Match each of the system requirements against the corresponding method and equipment capability for Parts A and B, part by part and item by item, and enter "Y" for yes, "N" for no, or "?" for undecided, in each of the columns marked "X" below, to indicate your conclusions. Delay comparison of method and equipment [illegible] current resources in Part C until the necessary information has been obtained from the manufacturers or suppliers for the classes marked "Y" or "?" in the OVER ALL CONCLUSIONS for Parts A and B. However, be sure to complete the two availability factors at the bottom of Part C.

CHARACTERISTIC OR FACTOR		METHOD AND EQUIPMENT CLASSES																			
		CLUE-WORD EXTRACT CARD	PERMUTED INDEX	COLUMNAR CARD	DUAL DICTIONARY	EDGE NOTCHED CARD	OPTICAL COINCIDENCE	EAM PUNCHED CARD	MISC. CARD SELECTORS	MICROFILM JACKET	MICROFICHE	MICROFILM STRIP	MICROFILM ROLL MECHANIZED	MICROFILM ROLL PHOTO-OPTICAL	MICROFILM CHIP AUTOMATED	MICROFILM EAM PUNCH CARD	MICROFILM NOTCHED CARD	MICROFILM SUPERMINIATURE	VIDEO TAPE	COMPUTER ALL	COMPUTER MASS MEMORY
Part A INPUT AND STORAGE	PHYSICAL																				
	FILE SIZE																				
	INTELLECTUAL																				
	SOURCE																				
	CHANGE																				
Part B RETRIEVAL AND PRESENTATION	ACTIVITY																				
	INTELLECTUAL																				
	OUTPUT PHYSICAL																				
	SERVICE																				
	QUALITY																				
OVER-ALL CONCLUSIONS - Parts A & B																					
Part C RESOURCES (Current)	PERSONNEL COSTS																				
	OTHER COSTS																				
	PERSONNEL AVAILABILITY																				
	OTHER AVAILABILITY																				

Figure 42

tion sheet and the nonconventional methods and equipment guide should be completed only after the necessary information has been obtained from the manufacturers and suppliers or other sources for the classes marked "Y" or "?" on the "Overall Conclusions—Parts A and B" line of the reconciliation sheet.

When these analyses are concluded, the person conducting the study should be ready to submit his findings, conclusions, and recommendations to management. The best solution to the information problem in many instances lies in a combination of methods and equipment—some of which may be new and some of which may be old. It is not only prudent and practical to retain those features of the old system that the users prefer, but also of considerable help in gaining acceptance of the new system.

Other records management handbooks that should be helpful in conducting this phase of the study are *Information Retrieval Systems*, a description of 50 operating information retrieval systems in Government and private industry; *Microform Retrieval Equipment Guide*, which describes the capabilities, characteristics, and costs of microfilm readers and reader printers; and, the *Source Data Automation Equipment Guide*, which explains the various techniques and equipment for capturing or converting data to machine language for automated processing.

IX. DESIGNING A COORDINATE INDEX

Most modern information retrieval systems employ some form of coordinate indexing. This chapter is mainly concerned with designing coordinate indexes employing manual indexing and used for retrieval of documents on the basis of their subject matter content. However, most of the guidelines also apply to designing systems used for conducting searches to identify people, places, or things on the basis of their characteristics, features, or attributes. The objective of this chapter is to provide guidance on the subject of designing a coordinate index and highlight the main considerations.

Economics of Coordinate Indexes

Investment in Input Versus Output. In a conventional system where the high cost of retrieving documents is mainly attributable to the inherent problems and limitations of conventional methods and equipment, the chances are that too little is now being invested in the input. While increases in indexing (input) effort will have a substantial effect initially on reducing retrieval (output) costs, the return is diminishing. A point is ultimately reached where further savings in output is possible only at a great additional investment in input, thus making the total cost per retrieval action higher than for a conventional system.

The lowest overall cost in any given situation can be achieved only by a proper apportionment of investment between input and output. Because usually far more information is entered into a system than will ever be retrieved, it is often better to forego some of the refinements in input, such as sophisticated linguistical controls, in favor of doing a little more work at the output stage, such as screening the search results. Figure 43 illustrates a range of input-output cost relationships that a systems design should consider in determining the maximum cost-benefit for a particular system.

Input Costs. In coordinate indexing systems, the main input costs are labor. If the system employs manual indexing techniques and is used for retrieval of documents on the basis of subject topics, the input effort is largely intellectual—man-hour requirements for analyzing incoming documents, and assigning index terms. If the system employs automatic indexing techniques or is used for identifying people, places, or things on the basis of their characteristics, features, or attributes, the major costs are for clerks and machine operators—man-hours for entering the information into the system. In both instances, system design and application of source data automation (SDA) techniques play a vital role in controlling input costs.

Effort Versus Results. It is important to recognize that in information retrieval, the total effort put into the system is subject to the laws of diminishing returns. No matter how much effort is put into collecting, organizing, and processing the information, the system itself will never be able to satisfy all the users' needs. There will always be instances where it may be more practical to rely on special handling, for example, consulting experts or other information sources or services for assistance.

Steps in Developing a Coordinate Index

While the methods used in developing a coordinate index will vary in accordance with the time available, the complexity of the situation, and other factors, there are certain essential steps. The sequence of the steps may vary from that shown below, and it is usually desirable to undertake some of these steps simultaneously:

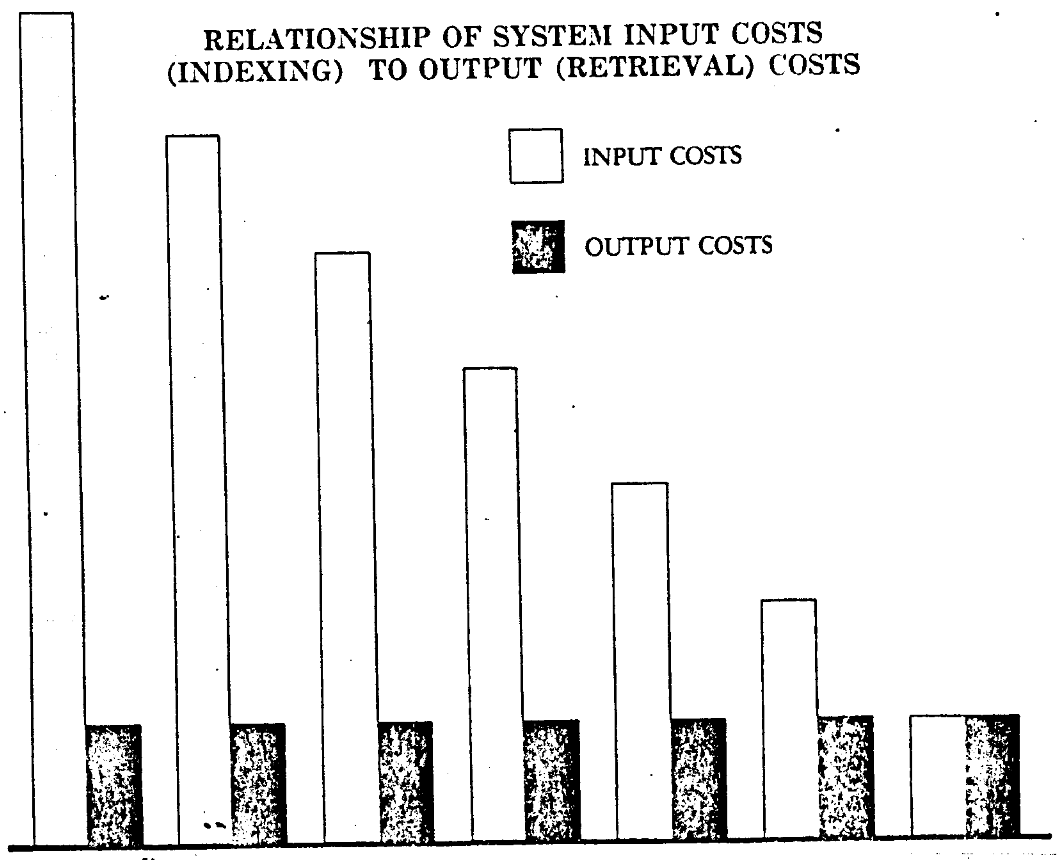
1. Review existing vocabularies.
2. Sample the documents.
3. Sample present searches.
4. Draft preliminary vocabulary.
5. Set up temporary index file.
6. Test and refine vocabulary.
7. Prepare the index manual.

Step 1: Review Existing Vocabularies. When developing a coordinate index one must be careful to select indexing terms on the basis of their significance in the subject matter field involved and their usefulness in conveying needed concepts. The review of existing vocabularies should include not only the formalized lists of descriptive terms, but also any other items that contain terms peculiar to the user group. Consequently, these sources of vocabulary material should be reviewed:

- Agency subject—classified outlines, subject indexes, or similar items.
- Organizational and functional charts and statements.
- Agency or installation annual reports and other publications describing the work of the organization.
- Laws, regulations, and directives.
- Index vocabularies in the same subject matter field developed by other Government agencies and private industry.

Step 2: Sample the Documents. A sampling should be made of the actual documents to be entered into the system in order to obtain a good idea of the range, scope, depth of coverage, and terminology used. If there are seasonal factors or other special circumstances, the selection method should be adjusted as necessary to obtain a representative sampling.

Step 3: Sample Present Searches. It is important to carefully study the present searches being made in order to obtain a good understanding of user language, preferences, and work habits.



Which of these relationships provides maximum cost-benefit at a particular installation?

Figure 43

number of indexing terms is far less than in keyword systems. Descriptors reduce the size of the index file and thereby save index storage and equipment costs and searching time. However, development and control of the descriptor indexing vocabulary requires professional know-how and trained, skilled indexers. Poor design, inconsistencies, and errors can reduce retrieval accuracy or even nullify the advantages of descriptor systems over keywords. Further, indexing is more time consuming and tedious than keyword systems since it involves subject analysis, looking up indexing terms in a glossary or thesaurus, and making decisions about which descriptor to use.

Keywords and descriptors may be used to considerable advantage in the same indexing system. When this is done descriptors usually serve as the main, official index vocabulary. Then, when indexing documents, the indexers are permitted to supplement the assigned descriptors with any keywords that they have learned from experience might be particularly helpful in retrieving the document later. Such keywords as trade names, popular jargon, and coined terms can thus be added to the index description of the document without disturbing the operation of the basic descriptor system.

Hard language or soft? A "hard" language or vocabulary is one in which the indexing terms are straightforward, well defined, and readily understood. Such terms as physical characteristics, quantitative measures, and geographical locations would produce a "hard" vocabulary by their very nature. Unfortunately, much of the language contained in documents to be retrieved by subjects is vague, imprecise, inconsistent, and abstract. Because such "soft" language invariably creates serious problems in indexing and in searching, one of the primary objectives in the construction of the index vocabulary should be to convert "soft" language to a more precise "hard" vocabulary of indexing terms.

Hardening of the vocabulary is accomplished by: (1) Careful treatment of synonyms and near-synonyms by deciding which term will be used and then cross-referencing the others to it. (Near-synonyms refer to words that have different dictionary definitions but which are frequently used interchangeably; for example, "mechanized" and "automated."); (2) Avoiding the use of terms that are not meaningful or which are so vague as to defy precise definition; (3) Developing clear definitions; (4) Using common standard technical terms, if they exist, in preference to trade names, lay terms, and short-lived coined or popular terms; (5) Using root words; that is, using the simple form of a word to cover all of its variations, sometimes referred to as "confounding"; for example, the word "extend" might include "extension," "extensive," "extended," and "extending"; (6) Using the noun form for all indexing terms; for example, use "pouring" instead of the verb "pour"; and (7) Using the plural rather than the singular form, except when referring to specific processes, properties, and conditions.

Step 5: Set Up a Temporary Index File. The index file is the medium upon which are recorded the indexing terms and other descriptive data used to identify individual documents. Columnar cards, optical coincidence cards, and computer magnetic tapes are some examples. Usually the temporary index file is of the same type that will be used for the permanent index record; however, in smaller files at least, simple handposted columnar cards may be used. Special measures should be taken, to the extent possible, to facilitate changes, additions and deletions in the temporary file. Steps should also be taken that will later permit incorporating the temporary file into the permanent file without having to redo the work. One of the ways to accomplish this is to prepare and retain individual paper tapes or EAM punched cards for the document as it is entered into the system during this period.

Arrangement of the coordinate index file. There are two basic ways for arranging the index file: (1) by document title or number; or (2) by indexing term or term number. (See figure 44.)

Document or conventional file arrangements consist of one index card or individual machine record for each document or item being indexed. All indexing terms and other descriptive data for a particular document or item are usually posted to its one index record. The index file is arranged by document title or number. The conventional file arrangement preserves the indexing of each individual document or item as an integral unit that can be helpful in analyzing the index file and correcting or changing index postings.

COORDINATE INDEX FILE ARRANGEMENTS

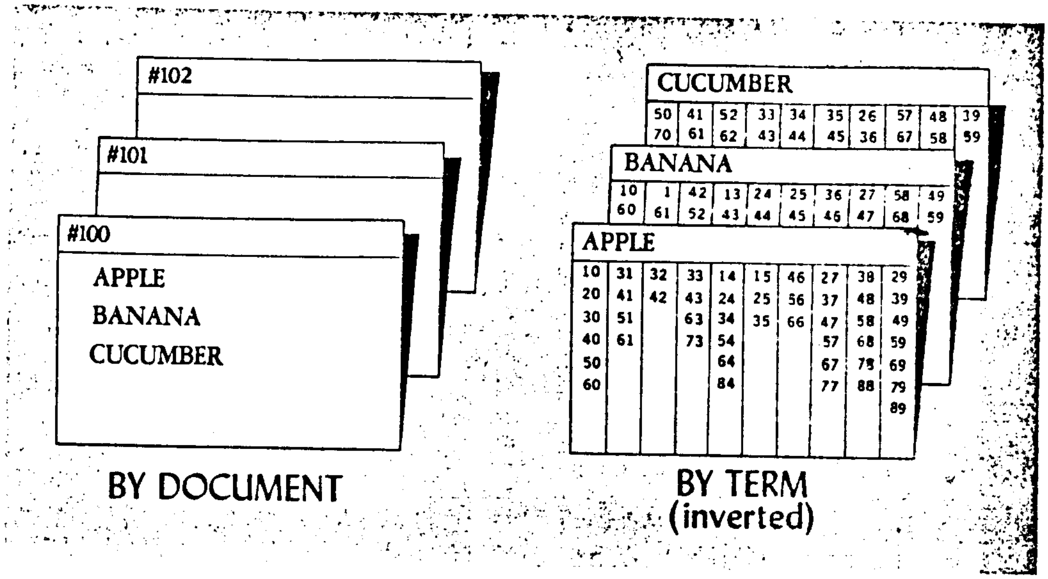


Figure 44

Conventional file arrangements also make it possible to have immediate knowledge of all the indexing terms assigned to the documents selected by the system during the search process, thus providing valuable clues as to their relevancy to the search question. However, such file arrangements require a large number of index records since there are usually several times as many documents or things to be indexed as there are indexing terms. For example, if there are 20,000 documents and 2,000 indexing terms, 20,000 index records would be needed. Conventionally arranged files require linear or serial searching of the file, which is usually more time consuming since every index record must be examined to determine [illegible] search question. For example, if only five indexing terms were used in the search, all 20,000 index records would still have to be examined.

Term or inverted file arrangements consist of one index card or individual machine record for each indexing term in the index vocabulary. The file is arranged by the indexing terms or term numbers. When the indexer has decided which indexing terms apply, the index records for those terms are selected and the document or item number is posted on each applicable index record. The inverted index file arrangement reduces to a minimum the number of individual index records that must be maintained. For example, if there are 20,000 documents or items and 2,000 indexing terms only 2,000 index records would be needed. Inverted file arrangements also greatly reduce the number of index records that must be examined, and thus also reduce the time required for the search process; for example, if five indexing terms were to be used in the search, only five index records would have to be examined. The major disadvantage of inverted file arrangements is that a search produces identifying numbers only, and it is therefore necessary to refer to another record to obtain descriptive information about the document and to determine its relevancy to the search question.

Term (inverted) and document (conventional) file arrangements are both sometimes used in the same system, particularly in those employing computers. The inverted file of indexing terms is maintained on-line to the computer to permit rapid searching of the entire file at one time. The search questions and the document numbers produced as a result of the search are then batched and periodically machine processed across a conventionally arranged magnetic tape index file containing the complete bibliographic information for each document in the system, possibly including an abstract. Thus the user can be furnished a printout showing the results of the search, including all the available bibliographic information.

Step 6: Test and Refine Vocabulary. This is the toughest and longest phase in the development of any coordinate index. The index vocabulary, like the retrieval system itself, must be tailored to the users' needs. One of the problems in testing and refining an index vocabulary is finding the right people to do the job. Ideally the individuals should have a thorough knowledge of the subject matter field plus training and experience in indexing. An acceptable substitute is the team approach in which professional people with knowledge of the subject are brought together. The testing and refining phase should cover at least 500 documents or a 6-month period, whichever occurs first. During this phase the temporary file should be used for actual searching, with tests made to determine the effectiveness of the vocabulary. Below are some of the things to look for and do in the testing and refining process. Figure 45 illustrates these points.

Broad or precise terms? The proper degree of indexing depth or specificity is governed by the size of the collection and user needs and can be arrived at only through a continuing analysis of these needs and system performance. In developing an index vocabulary, at the beginning one should lean toward use of broader terms in preference to the more specific terms until there is a proven need for the latter. The following are key criteria for determining how specific individual indexing terms should be:

- The terms ordinarily need be no more precise than those used in the material being indexed and by the users in their search requests. (Broad terms should ordinarily be used in areas of peripheral interest.)

- If the term receives heavy usage in indexing and heavy usage in searching and as a result more documents are retrieved than the users need or want, it probably should be replaced or supplemented by a more specific term. (It may still be necessary to retain the term in order to be able to conduct generic (general) searches.)

- If a term receives extremely light usage in indexing and searching, it probably should be dropped and included within the definition of another term, unless it is so unique or significant that it warrants retention as a separate term.

Single word or compound terms? In the early coordinate indexing systems individual terms consisted of a single word; however, it soon became apparent that there were times when two of the words should be joined. Words are joined together for one of the following reasons:

- They usually appear together in the document or form a single concept, for example, "North America," or "information retrieval."

- To provide specificity as in "metal tubing," "plastic tubing," etc.

- To prevent false retrieval caused by improper association of terms during the search process, for example, retrieval of a document about a "dog house" when the search concerned a "house dog."

While some combining of terms is necessary and beneficial, excessive or indiscriminate combining tends to defeat the basic purpose of coordinate indexing. It may result in loss of information at the time of retrieval and will increase the size of the index vocabulary.

General terms needed? Coordinate indexing, as explained earlier, is based on the principle of assigning numerous interdependent indexing terms which, when considered as a group, form a fairly complete description or, in effect, a limited abstract of the document. If the same indexing term is used for indexing documents that deal with a narrow aspect of a document and also for those that discuss the term in general, both types of documents will be retrieved if that term alone is used when conducting a search.

If searching for general documents under any particular indexing term is commonplace and re-

TESTING AND REFINING VOCABULARY TERMS

BROAD OR PRECISE TERMS?

BROAD	PRECISE
ANIMAL	Antelope Bear Cat Dog
FARMING	Cultivate Fertilize Irrigate Pollinate

SINGLE OR MULTIPLE WORDS?

SINGLE	MULTIPLE
New York	New York
Management Paperwork	Paperwork Management
Dog House	Dog House House Dog

GENERAL TERMS NEEDED?

SPEED
STOPOVER
SUPERCARGO
TRAILER
TRUCK
*TRUCKS, GENERAL
UNDERCARRIAGE
URBAN
UTILITY
WAGE
WEIGHT
WILDCAT STRIKE

Figure 45

sults in the retrieval of a large number of unwanted documents, some adjustment to the system may be necessary. One way would be to set up two indexing records for the term, one to be used when a document represents a general discussion of the terms and the other when the term is used in combination with other terms.

There are numerous other techniques for accomplishing this adjustment, including placing an asterisk beside the document number whenever it represents a general discussion of the term. However, as explained earlier, it is sometimes more practical to do a little extra screening of the output for the purpose of deleting unwanted documents than it is to try to improve system performance through additional refinements to the input process.

Step 7: Prepare the Indexing Manual. Even the simplest coordinate index system needs a manual. To make certain that the vocabulary is used as intended, it is necessary to put in writing the indexing rules, term definitions, and cross-references and to include appendixes of special reference aids needed. Indexing manuals go by many names, but all have one thing in common—they are the main control device of the coordinate index system. To the indexer, the manual is the system's "bible"; to the searcher and the user, it is an essential reference tool.

The index manual should serve as a translating tool for reconciling differences in the terms used by the authors and the users as well as to bridge the gap between the indexers and the searchers. This is accomplished by including all likely terms in the alphabetical listing of indexing terms and cross-referencing them to the equivalent terms used in the system.

It may be possible in a very small system to get by with a simple glossary, authority list, or dictionary of terms that includes definitions, where needed, and cross-references for synonyms. In the larger systems, where the indexing terms number in the hundreds or thousands, it becomes essential to know and display the relationships among the indexing terms—upward, downward, and horizontal. To answer this need, thesaurus-type indexing manuals are now in common use.

Construction of the thesaurus of indexing terms. Figure 46 shows a sample page from a thesaurus. The following is an explanation of the various headings:

Main index terms. These are the actual terms used for indexing documents. These same terms appear in the index file and constitute the index vocabulary of the system. Indexing terms consisting of two or more words should usually be listed by direct entry in their natural order; for example, RECORDS MANAGEMENT, not MANAGEMENT, RECORDS. In order to distinguish the various meanings of homographs, such qualifying expressions as TANKS (WEAPON) and TANKS (CONTAINER) may be used, in which event the qualifying expression becomes a part of the indexing.

Scope note. A short explanation used when needed to convey the meaning of an indexing term. A precise dictionary definition should not be attempted. The scope note merely indicates how the subject index term should be used and is not part of the subject index term:

> COMBUSTION CHAMBER GASES. The gases in a combustion chamber before or after ignition; for studies of gases ejected from the combustion chamber, see **EXHAUST GASES**.

Use reference (USE). The USE reference is intended to lead users of a thesaurus to appropriate subject index terms and should be employed to refer from a term that is not selected to one that is; for example:

1. To indicate a preferred synonym:

   SECONDARY BATTERIES *USE* STORAGE BATTERIES

2. To refer from a specific term to a more general term that has been selected to represent the specific concept:

   PLANT WAXES *USE* WAXES

   SAND BLASTING *USE* ABRASIVE BLASTING

3. To indicate a preference between spelling variations or to expand or explain abbreviations:

   INFLAMMABILITY *USE* FLAMMABILITY

   PENTAERYTHRITOL TETRANITRATE *USE* PETN

   EEG *USE* ELECTROENCEPHALOGRAMS

4. To express concepts that can be considered synonyms for purposes of indexing and retrieval:

SAMPLE PAGE FROM THESAURUS OF INDEXING TERMS

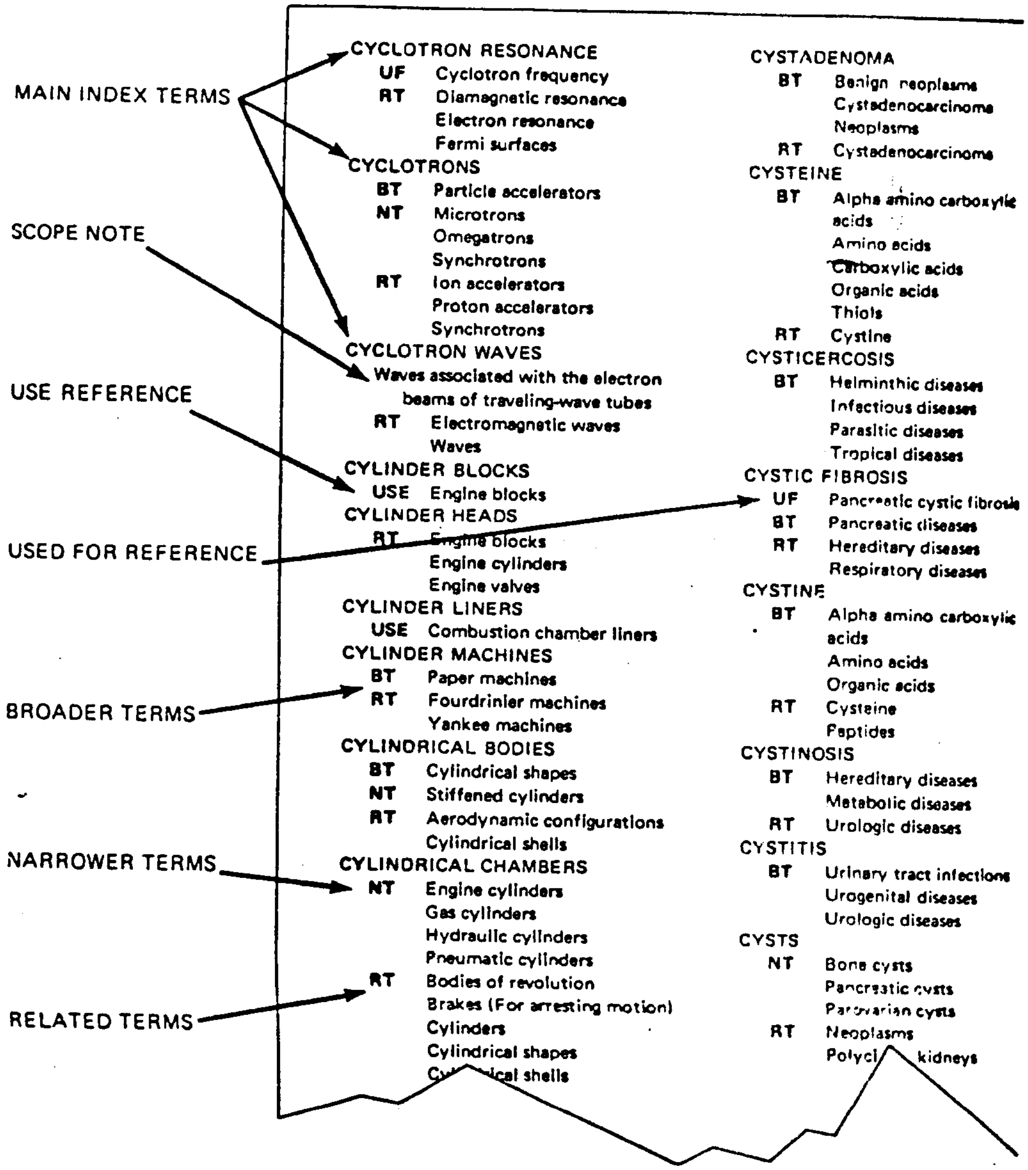


Figure 46

HEREDITY *USE* GENETICS

5. To bring together different viewpoints of a conceptual continuum:

SMOOTHNESS *USE* ROUGHNESS

6. To explain variations in word order:

TABLES (MATHEMATICS) *USE* MATHEMATICAL TABLES

7. To show how two or more index terms can be assigned to express a word not included in the index vocabulary:

HEN *USE* FEMALE AND CHICKEN

Used for reference (UF). The UF reference is the reciprocal of the USE reference. It should be used because it is essential for recordkeeping.

STORAGE BATTERIES *UF* SECONDARY BATTERIES

Broader terms (BT) and narrower terms (NT). The BROADER TERM (BT) and NARROWER TERM (NT) cross-references are employed to indicate class relationships that may exist among subject index terms. The reference is used to refer from a term symbolizing a concept class to all terms symbolizing concepts that are members of that class. The reference is used to refer from a term representing a member of a class of concepts to the term in the thesaurus representing that class. Whenever either of these cross-references is used, the reciprocal reference is also entered:

STEELS *BT* IRON ALLOYS

IRON ALLOYS *NT* STEELS

Related term (RT). The RT cross-reference is employed to refer from a subject index term to any other terms that are closely related conceptually but not hierarchically. For recordkeeping purposes, RELATED TERM references should always be entered reciprocally:

ORES *RT* MINERALS

MINERALS *RT* ORES

Hierarchical reference aids. Earlier discussions have disclosed the problems and limitations of trying to organize large bodies of complex, changing material in a hierarchical classification basis for retrieval by subjects. However, the change to a coordinate index does not eliminate the need or desirability for being able to determine hierarchical "family tree" relationships among terms. It is a natural inclination of many people to classify and organize information and items hierarchically because this is the method most familiar to them; consequently, they prefer that the reference aid be organized in this manner. Hierarchical classification schemes have their own "built-in" logic that helps the system designer, the indexer, the searcher, and the user get an overall picture of the coverage and scope of the collection and the depth of indexing.

Some manuals, therefore, also include hierarchical finding aids in which terms appearing in the straight alphabetical listing are arranged hierarchically. (See figure 47). Since these finding aids in no way change the structure of the actual vocabulary or the arrangement of the index file, several different ones can be developed, if needed, to reflect the preferences and needs of various types of user groups.

Staffing

It is futile to attempt to establish an information retrieval system without competent personnel. Otherwise, the best designed system will not be effective and a weak system may not survive long enough to give the designer an opportunity to correct the design deficiencies. A key question in planning personnel needs is: "Should subject matter specialists or professional indexers be secured?" In systems for retrieval by subjects, the need for a thorough knowledge of the subject field and the art of indexing are probably of equal importance. If a choice must be made between candidates who have only one of these skills, it is usually better to select the person who has a thorough knowledge of the subject field and then train him to be an indexer. An exception to this would be a situation where the system is used for storage and retrieval of routine general material such as newspaper clippings, in which case it should be possible, with the aid of a good indexing manual, to train any reasonably intelligent person to do the job.

Current Awareness Services

In addition to retrieving documents or data upon demand (retrospective searching), three other services that are sometimes incorporated in a coordinate indexing system are issuance of document announcement bulletins, abstracting, and selective dissemination of information. These types of current awareness services are designed to inform potential users of information about the availability and contents of recently received documents.

Announcement Bulletins. Printed periodical announcement bulletins are issued in situations where there are a large number of user groups. They list in numerical sequence descriptive infor-

EXAMPLES OF HIERARCHICAL FINDING AIDS FOR A THESAURUS

HEAT and THERMODYNAMICS

- **Heating Plants**
  - RADIANT HEATING
  - SOLAR FURNACES
- **Instrumentation**
  - CALORIMETERS
  - THERMOMETERS
- **Physical Reaction**
  - HEAT TRANSFER
  - THERMAL EXPANSION

RADIOACTIVITY

- **Decay**
  - ALPHA DECAY
  - BETA DECAY
- **Hazards**
  - CONTAMINATION
  - FALLOUT
- **Chemical Analysis**
  - CHROMATOGRAPHIC
  - COLOREMETRIC

Figure 47

mation (abstracts) on newly acquisitioned documents at an information center and usually include subject or author indexes to aid in finding particular documents listed. See figure 48 for a sample of such a bulletin. Even in small information facilities, where formal published bulletins are not warranted, some method is needed to keep users informed of the availability of new accessions.

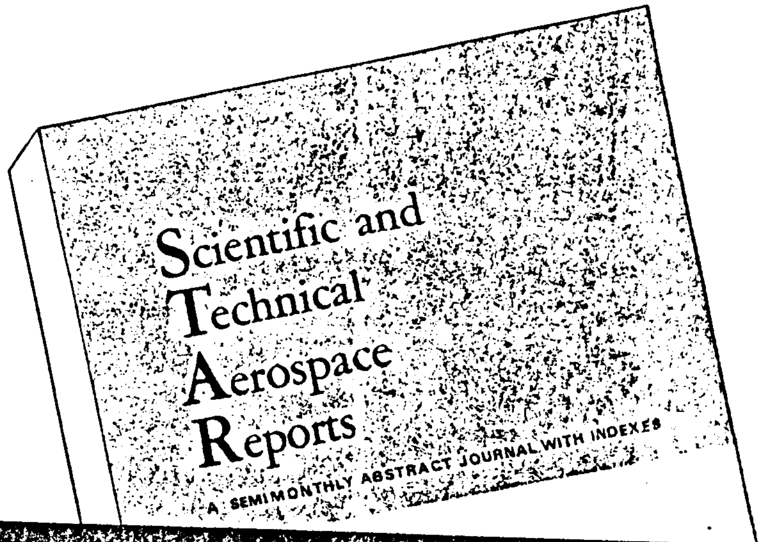
Abstracts. Because abstracting is expensive, its use should normally be restricted to situations where the documents receive widespread distribution or use. Many documents received from outside sources may already include abstracts that may be incorporated in the system at little or no expense. Most abstracts are prepared by professional indexers and editors; however, there is a growing tendency to require the authors to prepare the abstracts, a practice which in a few instances has met with failure, yet in other instances has been successful. Figure 48 also includes samples of abstracts of newly accessioned documents at the Scientific and Technical Information Facility of National Aeronautics and Space Administration. Both author- and indexer-prepared abstracts are included in this system.

Author abstracting should be given serious consideration in such fields as law, medicine, and others where case histories and decisions need to be disseminated and recorded for future study. Some professional assistance and editing may still be required, of course. Perhaps one of the greatest values of abstracts lies in their potential future use as input for automatic indexing and machine retrieval of documents.

Selective dissemination of information (SDI). As explained in chapter V, SDI involves notifying the user (or user groups) individually each time a document is received which is of the type the user has indicated might be of interest to him. To accomplish this, each user's interest profile is developed, with his help, and often maintained on computer tape. The computer compares the indexing description of each new incoming document against the user interest profiles; if they match, the computer prepares a notice that is sent to the user. The notice usually contains a description of the document and the user is given the opportunity to borrow or acquire a copy.

Effect of current awareness services. The use of announcement bulletins, selective dissemi-

SAMPLE OF A PUBLISHED ANNOUNCEMENT BULLETIN



16-13 GEOPHYSICS

N71-28798# Applied Physics Lab., Johns Hopkins Univ., Silver Spring, Md.
ALGORITHM TO COMPUTE TROPOSPHERIC REFRACTION EFFECTS ON RANGE MEASUREMENTS
S. M. Yionoulis Jul. 1970 13 p refs
(Contract N00017-62-C-0604)
(AD-721333; APL-TG-1125) Avail: NTIS CSCL 4/1

H. S. Hopfield has modeled a two-quartic tropospheric refractivity profile for correcting satellite range and range rate data which treats the dry and wet components of the tropospheric refractivity separately. The expression given for computing the contributions to range data is sensitive to rounding errors at high elevations even when evaluated in double precision. Alternate forms of the solution are presented here which eliminate the problem of rounding errors and the need for double precision computation. They allow the user to benefit from the full accuracy of Mrs. Hopfields model for all elevations. Author (GRA)

N71-28853# National Oceanic and Atmospheric Administration, Washington, D C Federal Coordinator for Meteorological Services and Supporting Research
FEDERAL PLAN FOR AIR POLLUTION CONTROL METEOROLOGICAL SERVICE: TO SUPPORT FEDERAL, STATE, AND LOCAL AIR POLLUTION CONTROL AGENCIES
Jan. 1971 27 p
(COM-71-00200, NOAA-71012801) Avail. NTIS CSCL 13B

The needs of federal, state, and local air pollution control agencies for specialized meteorological support are investigated Arency responsibilities, service concepts, and a 5-year program directed toward providing improved services to these users are described Author (GRA)

N71-28866 National Lending Library for Science and Technology, Boston Spa (England).
FLUCTUATIONS OF THE TOTAL OZONE CONTENT OF THE ATMOSPHERE IN CONNECTION WITH STRATOSPHERIC WARMINGS
W Hoebbel 24 Jun 1970 9 p refs Transl into ENGLISH from Met. Dienst DDR, 75 Jahre Met Obs Potsdam 1892 - 1967 (Potsdam), 1969 p 108 - 111
(NLL-M-9270-(5828 4F)) Avail Natl Lending Library, Boston Spa, Engl. 1 NLL photocopy coupon

Monthly means of total ozone content averaged over all stations in the Northern Hemisphere were always greater in magnitude in years in which there was a late final warming than in years with an early warming A relationship between motion direction of sudden stratospheric temperature change regions and wind direction at the equatorial region and a high total ozone content over the Northern Hemisphere is established G G.

N71-28867*# Translation Consultants, Ltd., Arlington, Va.
CHANGE IN THE ELASTIC PARAMETERS AND STRENGTH OF ROCKS UNDER PRESSURE [OB IZMENENII UPRUGIKH PARAMETROV I PROCHNOSTI GORNYKH POROD POD DEYSTVIYEM DAULENIYA]
Z I Stakhovskaya Washington NASA Jun. 1971 8 p refs Transl into ENGLISH from the publ. 'Problemy Mekhaniki Gornykh Porod. Vsesoyuznoy Nauchnoy Konferentsii, 1st' Alma-Ata 1965 p 394 - 398
(Contract NASw-2038)
(NASA-TT-F-13653) Avail NTIS CSCL 08G

Models are used to demonstrate the effect porosity in rocks has on change in elastic parameters when rocks are subjected to high uniform pressures It is shown that the change can be found by knowing the relationship between areas of contacts and units strain, and that this latter relationship can be found by using certain formulas and curves, both of which are included Author

Figure 48

nation, abstracts, and other current awareness services are a valuable means for communicating new ideas and information and can be instrumental in reducing duplication of effort. To the designer of the information retrieval system, the incorporation of such current awareness techniques in the system is of major concern, since these techniques can substantially reduce the retrospective searching workload. The more that is done in the way of current awareness, the less searching that is required, usually. Also, unless users are kept informed and given a simple, easy method for obtaining current information, they are likely to turn to their colleagues for the needed information or to their personal files rather than to use the information facility.

Quality Control

It is necessary to achieve acceptable quality in every retrieval system, but the art is fraught with too many problems to ever be perfect. The term "quality" as used here refers to the percent of recall and precision and the absence of errors and inconsistencies.

Recall. Recall represents the percent of pertinent documents known to be in the collection that are retrieved in response to a search question. If a system has high recall, it means that only a few pertinent documents are being missed or overlooked when subject searches are made. Low recall, on the other hand, means that a substantial percentage of pertinent documents are not being retrieved.

Precision (or Relevance). Precision represents the percent of documents retrieved during a subject search that are relevant to the search question. If a system has high precision, it means that the users find that only a few irrelevant documents are being retrieved. Low precision, on the other hand, means that a large percentage of the documents retrieved are not pertinent to the search question.

Errors and Consistency. Indexing errors and lack of consistency are another major cause of indexing systems failures.

Setting Quality Standards

It is just about impossible to achieve 100 percent in both recall and precision. Improvements in recall tend to decrease precision and vice versa. However, system performance can be improved by various means. The all-important thing to remember is that management should decide what standards it wants the system to achieve; i.e., high recall and low precision, low recall and high precision, or somewhere in between. The higher standards require more costly controls, and management must weigh the value of different levels of performance in the light of the costs of achieving these levels.

Achieving Higher Recall Performance. These are the various ways that recall performance can be increased:

Harden vocabulary by careful treatment of synonyms, avoiding the use of vague terms, developing clear definitions, using standard technical terms in preference to popular jargon, and using root words to cover all variations of a term.

Use broader terms in both the vocabulary and in the assignment of terms to individual documents.

Assign more terms per document so that those topics or concepts only slightly involved are also included in the index descriptions.

More exhaustive searching by broadening the search and improving the search strategy.

Improving Precision Performance. These are some actions that can improve precision performance:

Increase vocabulary specificity by working closely with the users to develop terms that will express the needed information more precisely.

Add weights to each term assigned to the document. For example, a "1" following an index term "corrosion" might mean that the document contains information of major importance on that topic; a "2" might signify moderate importance; and a "3," minor importance. Or an asterisk could be placed in front of a term whenever it is of major importance.

Increase search specificity by having the searchers work more closely with the users in negotiation of the search in order to select more precise terms.

Reducing Errors and Increasing Consistency. These are various ways to reduce errors and increase consistency:

Training. Develop a systematic plan for training new indexers and searchers and refresher courses for experienced employees.

Prescriptive indexing. Wherever possible, prescribe in the indexing manual which term will be used in situations where there are various possibilities, instead of leaving the choice to the indexer.

Indexing and searching aids. Develop hierarchical or other "lead-in" vocabularies as an appendix to the indexing manual; also develop written rules for search strategy.

Personnel rotation. Rotate personnel between indexing and searching; also consider rotating personnel within the indexing group.

Spot checks. Use spot checks or random sampling quality control techniques. (Complete review of all indexing work would normally be too costly and usually does not completely solve the problem anyway.)

Conclusion

This chapter makes it quite clear that designing and operating a coordinate index is a formidable task. However, coordinate indexing systems offer the most powerful technique yet developed for manually organizing information and retrieving it and are essential to meet many of today's complex information needs. The only other possibility is the automatic indexing and searching system described in chapter V, which is, in effect, a form of coordinate indexing. The theory of automatic indexing is about as old as coordinate indexing; however, its development and growth have been much slower, largely due to the higher initial and input costs and the shortage of people having experience in the field.

It should also be understood that there is no such thing as a finished design for a coordinate index system. Most systems will require substantial revisions in a year or two after being established, and major revisions will occur approximately every five years. Consequently, a systems designer intimately familiar with the system should be available periodically to evaluate the performance of the system and develop plans for making the changes.

APPENDIX A

INFORMATION RETRIEVAL
NONCONVENTIONAL METHODS AND EQUIPMENT GUIDE

PART A - INPUT AND STORAGE

<table>
<tr><th colspan="2" rowspan="3">CHARACTERISTIC OR FEATURE</th><th colspan="4">CLUE-WORD EXTRACT CARD</th><th colspan="4">PERMUTED INDEX</th><th colspan="4">COLUMNAR CARD</th><th colspan="4">DUAL DICTIONARY</th></tr>
<tr><th>DR</th><th>DS</th><th>URS</th><th>DFR</th><th>DR</th><th>DS</th><th>URS</th><th>DFR</th><th>DR</th><th>DS</th><th>URS</th><th>DFR</th><th>DR</th><th>DS</th><th>URS</th><th>DFR</th></tr>
<tr><td>X</td><td></td><td></td><td></td><td>X</td><td></td><td></td><td></td><td>X</td><td></td><td></td><td></td><td>X</td><td></td><td></td><td></td></tr>
<tr><td rowspan="4">PHYSICAL</td><td>MAXIMUM DOCUMENT SIZE</td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>IDEAL DOCUMENT (OR DATA ELEMENT) LENGTH (3)</td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>MAXIMUM DOCUMENT (OR DATA ELEMENT) LENGTH (3)</td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>RESTRICTIONS ON FORM OF INFO THAT MAY BE STORED</td><td colspan="4">None</td><td colspan="4">Written info only</td><td colspan="4">Index terms and related document numbers only</td><td colspan="4">Index terms and related document numbers only</td></tr>
<tr><td rowspan="3">FILE SIZE</td><td>IDEAL TOTAL NUMBER OF DOCUMENTS (OR DATA ELEMENTS) (3)</td><td colspan="4">INA</td><td colspan="4">500 to 2000 per listing</td><td colspan="4">500 to 5000</td><td colspan="4">Under 1000</td></tr>
<tr><td>IDEAL TOTAL VOLUME OF INFO</td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>LIMITATIONS ON FURTHER EXPANSION</td><td colspan="4">None except possibly space</td><td colspan="4">Requires multiple sets or reissue</td><td colspan="4">Ease and speed of retrieval</td><td colspan="4">Ease and speed of retrieval</td></tr>
<tr><td rowspan="6">INTELLECTUAL</td><td>IDEAL AMOUNT OF DESCRIPTIVE DATA PER DOCUMENT (OR DATA ELEMENT) (3)</td><td colspan="4">Less than half a page</td><td colspan="4">Up to 80 characters</td><td colspan="4">3 to 6 index terms</td><td colspan="4">3 to 6 index terms</td></tr>
<tr><td>IDEAL TOTAL AMOUNT OF DESCRIPTIVE DATA PER SYSTEM</td><td colspan="4">INA</td><td colspan="4">40,000 to 160,000 per listing</td><td colspan="4">50 to 500 terms</td><td colspan="4">50 to 500 terms</td></tr>
<tr><td>CAPABILITY - EXPANSION OF DESCRIPTIVE OR OTHER DATA</td><td colspan="4">Excellent</td><td colspan="4">Very limited</td><td colspan="4">Excellent</td><td colspan="4">Excellent</td></tr>
<tr><td>CAPBILITY - REORGANIZATION OF DESCRIPTIVE OR OTHER DATA</td><td colspan="4">Excellent</td><td colspan="4">Not necessary</td><td colspan="4">Excellent</td><td colspan="4">Excellent</td></tr>
<tr><td>CAPABILITY - CHANGING DOCUMENT ARRANGEMENT</td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>SPECIAL SKILLS REQUIRED FOR INPUT PROCESSING (1)</td><td colspan="4">None</td><td colspan="4">Keypunch and ADP Operators</td><td colspan="4">None</td><td colspan="4">None</td></tr>
<tr><td rowspan="3">SOURCE</td><td>CAPABILITY - ACCEPTING DATA IN MACHINE LANGUAGE</td><td colspan="4">None</td><td colspan="4">Excellent</td><td colspan="4">Only if maintained on computer</td><td colspan="4">Excellent</td></tr>
<tr><td>CAPABILITY - PRODUCING SYSTEM BY MEANS OF A COMPUTER</td><td colspan="4">Good</td><td colspan="4">Essential</td><td colspan="4">Good, but requires reissue to update</td><td colspan="4">Essential</td></tr>
<tr><td>NEED FOR STANDARDIZED FORM AND FORMAT FOR INPUT</td><td colspan="4">No special requirements</td><td colspan="4">Essential</td><td colspan="4">No special requirements</td><td colspan="4">No special requirements</td></tr>
<tr><td rowspan="4">CHANGE</td><td>CAPABILITY - ADDING TO DESCRIPTIVE OR OTHER STORED DATA</td><td colspan="4">Excellent</td><td colspan="4">only if reissued</td><td colspan="4">Excellent</td><td colspan="4">Only if reissued</td></tr>
<tr><td>CAPABILITY - CHANGING DESCRIPTIVE OR OTHER STORED DATA</td><td colspan="4">Good</td><td colspan="4">Only if reissued</td><td colspan="4">Good, but time consuming</td><td colspan="4">Only if reissued</td></tr>
<tr><td>CAPABILITY - ADDING TO CONTENTS OF STORED DOCUMENTS</td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td>CAPABILITY - CHANGING CONTENTS OF STORED DOCUMENTS</td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td></tr>
</table>

ABBREVIATIONS AND SYMBOLS

INA Information not available
D & D See Definitions & Descriptions

1 Refers to machine or equipment skills, only. In addition, subject matter knowledge on a par with that of users may be needed.

2 Refers to general accuracy when that class of method or equipment is used. Does not take into consideration human factors affecting accuracy or quality of the results.

3 Quantities shown for EAM Punched Card, Computer, and Computer Mass Memory refer to data elements rather than the entire document or record.

<table>
<tr><th colspan="4">EDGE NOTCHED CARD</th><th colspan="4">OPTICAL COINCIDENCE</th><th colspan="4">EAM PUNCHED CARD (3)</th><th colspan="4">MISC. CARD SELECTORS</th><th colspan="4">MICROFILM JACKET</th><th colspan="4">MICROFICHE</th></tr>
<tr><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td></tr>
<tr><td>X</td><td></td><td></td><td>-X</td><td>X</td><td></td><td></td><td>-X</td><td>X</td><td></td><td></td><td>-X</td><td></td><td>X</td><td></td><td></td><td></td><td>X</td><td></td><td></td><td></td><td>X</td><td></td><td></td></tr>
<tr><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4">5×8, no limit if on microfilm</td><td colspan="4">14" wide, any length or 40×60"</td><td colspan="4">Normally 8½×11"</td></tr>
<tr><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4">Same as that for record medium</td><td colspan="4">9 to 60 pages</td><td colspan="4">9 to 60 pages</td></tr>
<tr><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4">Same as that for record medium</td><td colspan="4">Impractical for over 120 pages</td><td colspan="4">Impractical for over 120 pages</td></tr>
<tr><td colspan="4">None</td><td colspan="4">Index terms and related document numbers only</td><td colspan="4">Alphanumeric characters and symbols</td><td colspan="4">Same as that for record medium</td><td colspan="4">None</td><td colspan="4">None</td></tr>
<tr><td colspan="4">1000 to 5000</td><td colspan="4">1000 to 10,000</td><td colspan="4">DR 1000 to 20,000
DFR 1000 and up</td><td colspan="4">1000 cards and up</td><td colspan="4">5000 and up</td><td colspan="4">10,000 and up</td></tr>
<tr><td colspan="4"></td><td colspan="4"></td><td colspan="4">DR-N/A, DFR up to 1,500,000</td><td colspan="4">1000 pages or sheets and up</td><td colspan="4">50,000 pages and up</td><td colspan="4">100,000 pages and up</td></tr>
<tr><td colspan="4">Ease and speed of retrieval</td><td colspan="4">Multiple setslowers search speed</td><td colspan="4">Lower retrieval speed</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">None</td></tr>
<tr><td colspan="4">3-6 items of coded data, or half page of other info</td><td colspan="4">3 to 10 index terms</td><td colspan="4">Under 80 characters</td><td colspan="4">Varies, usually under 15 digits</td><td colspan="4">Under 45, but OK up to 180</td><td colspan="4">Under 45, but OK up to 180</td></tr>
<tr><td colspan="4">No ideal amount</td><td colspan="4">50-1000 index terms</td><td colspan="4">80,000 up</td><td colspan="4">No ideal amount</td><td colspan="4">45-180 × the no. of documents</td><td colspan="4">45-180 × the no. of documents</td></tr>
<tr><td colspan="4">Limited</td><td colspan="4">Satisfactory</td><td colspan="4">Excellent, if the space permits</td><td colspan="4">None, as far as the selector is concerned</td><td colspan="4">OK up to 180 Char.</td><td colspan="4">OK up to 180 Char.</td></tr>
<tr><td colspan="4">Limited</td><td colspan="4">Excellent</td><td colspan="4">Excellent</td><td colspan="4">None, as far as the selector is concern</td><td colspan="4">Possible, but time consuming</td><td colspan="4">Usually requires refilming</td></tr>
<tr><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4">Not necessary due to random fililing</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td></tr>
<tr><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">Keypunch and ADP operators</td><td colspan="4">None</td><td colspan="4">Microfilming</td><td colspan="4">Microfilming</td></tr>
<tr><td colspan="4">None</td><td colspan="4">Some systems are excellent</td><td colspan="4">Excellent</td><td colspan="4">Usually none</td><td colspan="4">None</td><td colspan="4">None</td></tr>
<tr><td colspan="4">None</td><td colspan="4">Some systems are excellent</td><td colspan="4">Excellent</td><td colspan="4">Same as that for record medium</td><td colspan="4">Excellent</td><td colspan="4">Presently limited</td></tr>
<tr><td colspan="4">Essential</td><td colspan="4">Essential</td><td colspan="4">Essential</td><td colspan="4">Essential</td><td colspan="4">No special requirements</td><td colspan="4">Essential</td></tr>
<tr><td colspan="4">Only if the space permits</td><td colspan="4">Excellent</td><td colspan="4">Excellent, if the space permits</td><td colspan="4">Same as that for record medium</td><td colspan="4">Good</td><td colspan="4">Limited</td></tr>
<tr><td colspan="4">Very limited</td><td colspan="4">Very limited</td><td colspan="4">Excellent, but time consuming</td><td colspan="4">Same as that for record medium</td><td colspan="4">Limited</td><td colspan="4">Requires refilming or index, usually</td></tr>
<tr><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4">Same as that for record medium</td><td colspan="4">Excellent</td><td colspan="4">Requires refilming or index, usually</td></tr>
<tr><td colspan="4"></td><td colspan="4"></td><td colspan="4"></td><td colspan="4">Same as that for record medium</td><td colspan="4">Limited</td><td colspan="4">Requires filming or index, usually</td></tr>
</table>

PART A - INPUT AND STORAGE (Continued)

	CHARACTERISTIC OR FEATURE	MICROFILM STRIP	MICROFILM ROLL MECHANIZED	MICROFILM ROLL PHOTO-OPTICAL	MICROFILM CHIP AUTOMATED
		DR: ; DS: -X; URS: ; DFR: -X	DR: ; DS: X; URS: ; DFR: -X	DR: X; DS: X; URS: X; DFR: -X	DR: ; DS: X; URS: -X; DFR:
PHYSICAL	MAXIMUM DOCUMENT SIZE	14" wide, any length or 40 × 60"	14" wide any length	14" wide any length	8½ × 14" or 40 × 60"
PHYSICAL	IDEAL DOCUMENT (OR DATA ELEMENT) LENGTH (3)	DS - up to 40 pages DFR-25 to 50 lines	DS - one page DFR-50 lines	DR-any; DS URS 1-8 pages; DFR-25-50 lines	DS - 1 to 2 pages URS-1 to 8 pages
PHYSICAL	MAXIMUM DOCUMENT (OR DATA ELEMENT) LENGTH (3)	DS-none DFR-1000 lines	Not necessary if over 100 pages or 5000 lines	DR-none; DS URS/ DFR-not nec. if over 25 pages	Should not exceed above
PHYSICAL	RESTRICTIONS ON FORM OF INFO THAT MAY BE STORED	None	None	None	None
FILE SIZE	IDEAL TOTAL NUMBER OF DOCUMENTS (OR DATA ELEMENTS) (3)	1000 and up	10,000 and up	25,000 and up	DS-50,000 and up URS-100,000 and up
FILE SIZE	IDEAL TOTAL VOLUME OF INFO	DS - up to 40,000 DFR-25,000 lines and up	DS - 10,000 and up DFR - 250,000 lines and up	DS URS-25,000 and up DFR-million lines and up	DS-50,000 pages up URS-500,000 up
FILE SIZE	LIMITATIONS ON FURTHER EXPANSION	None	None	URS-speed and cost others- none	None
INTELLECTUAL	IDEAL AMOUNT OF DESCRIPTIVE DATA PER DOCUMENT (OR DATA ELEMENT) (3)	Under ten characters	Under ten characters	Varies-9 digits or 56 characters	DS-up to 18 digits URS-up to 100 Char.
INTELLECTUAL	IDEAL TOTAL AMOUNT OF DESCRIPTIVE DATA PER SYSTEM	10.000 char. and up	100,000 characters and up	200,300 digits or characters and up	DS-900,000 digits up URS-10 M and up
INTELLECTUAL	CAPABILITY - EXPANSION OF DESCRIPTIVE OR OTHER DATA	Limited and may require refilming	Limited and may require refilming	Requires refilming	DS-requires refilming URS-OK up to 100
INTELLECTUAL	CAPABILITY - REORGANIZATION OF DESCRIPTIVE OR OTHER DATA	Requires refilming	Requires refilming	Requires refilming	DS-requires refilming URS-excellent
INTELLECTUAL	CAPABILITY - CHANGING DOCUMENT ARRANGEMENT	Requires refilming	Requires refilming	Requires refilming	Good
INTELLECTUAL	SPECIAL SKILLS REQUIRED FOR INPUT PROCESSING (1)	Microfilming	Microfilming	Microfilming and ADP type skills	Microfilming and ADP type skills
SOURCE	CAPABILITY - ACCEPTING DATA IN MACHINE LANGUAGE	None	None	Limited	DS-none URS-excellent
SOURCE	CAPABILITY - PRODUCING SYSTEM BY MEANS OF A COMPUTER	Excellent	Excellent	16 MM, excellent	Partial
SOURCE	NEED FOR STANDARDIZED FORM AND FORMAT FOR INPUT	DS-none DFR-essential	Essential	Essential	Essential
CHANGE	CAPABILITY - ADDING TO DESCRIPTIVE OR OTHER STORED DATA	Requires slicing, refilming, or index	Requires splicing, refilming, or index	Requires splicing, refilming, or index	DS-requires refilming URS-excellent
CHANGE	CAPABILITY - CHANGING DESCRIPTIVE OR OTHER STORED DATA	See above	See above	See above	DS-requires refilming URS-excellent
CHANGE	CAPABILITY - ADDING TO CONTENTS OF STORED DOCUMENTS	See above	See above	See above	Satisfactory by adding new chips
CHANGE	CAPABILITY - CHANGING CONTENTS OF STORED DOCUMENTS	See above	See above	See above	Requires refilming

<table>
<tr><td colspan="4">MICROFILM EAM PUNCH CARD</td><td colspan="4">MICROFILM NOTCHED CARD</td><td colspan="4">MICROFILM SUPERMINIATURE</td><td colspan="4">VIDEO TAPE</td><td colspan="4">COMPUTER-ALL (3)</td><td colspan="4">COMPUTER MASS MEMORY (3)</td></tr>
<tr><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td><td>DR</td><td>DS</td><td>URS</td><td>DFR</td></tr>
<tr><td></td><td>X</td><td>X</td><td></td><td></td><td>X</td><td>X</td><td></td><td></td><td>X</td><td></td><td></td><td></td><td>X</td><td></td><td>-X</td><td>X</td><td></td><td></td><td>X</td><td>X</td><td></td><td></td><td>X</td></tr>
<tr><td colspan="4">14" wide, any length or 40 x60"</td><td colspan="4">14" wide, any length, or 40 x 60</td><td colspan="4">INA</td><td colspan="4">Up to 8½ x 14"</td><td colspan="4"></td><td colspan="4"></td></tr>
<tr><td colspan="4">1 to 8 pages</td><td colspan="4">1 to 8 pages</td><td colspan="4">INA</td><td colspan="4">DS-one page
DFR-25-50 lines</td><td colspan="4">Up to 100 characters</td><td colspan="4">Up to 100 characters</td></tr>
<tr><td colspan="4">Should not exceed above</td><td colspan="4">Should not exceed above</td><td colspan="4">INA</td><td colspan="4">Cost and efficiency loss</td><td colspan="4">None</td><td colspan="4">None</td></tr>
<tr><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">Numeric, alpha and special characters</td><td colspan="4">Numeric, alpha and special characters</td></tr>
<tr><td colspan="4">DS-10,000 and up
URS-1000-20,000</td><td colspan="4">1000 - 5000</td><td colspan="4">Undetermined</td><td colspan="4">200,000 and up</td><td colspan="4">20,000 and up
no fixed ideal</td><td colspan="4">Over 1,000,000 on line</td></tr>
<tr><td colspan="4">DS-10,000 pages up
URS-1K - 160,000 page</td><td colspan="4">1000-40,000 pages</td><td colspan="4">Undetermined</td><td colspan="4">DS-200,000 pages up
DFR-5million lines</td><td colspan="4">200,000 up
no fixed ideal</td><td colspan="4">Over 100,000,000 on line</td></tr>
<tr><td colspan="4">DS - none
URS-search time</td><td colspan="4">Speed</td><td colspan="4">Undetermined</td><td colspan="4">Speed and cost</td><td colspan="4">On line physical
Off line-time</td><td colspan="4">On line-physical
Off line-time</td></tr>
<tr><td colspan="4">Up to 58 characters</td><td colspan="4">3 - 6 coded items and half page of other</td><td colspan="4">Undetermined</td><td colspan="4">Up to 18 numeric or 12 alpha characters</td><td colspan="4">Up to 100% of the ideal document length</td><td colspan="4">Up to 100% of the ideal document length</td></tr>
<tr><td colspan="4">DS-58,000 char. up
URS-50,000 to 1M</td><td colspan="4">No ideal amount</td><td colspan="4">Undetermined</td><td colspan="4">2million characters up</td><td colspan="4">See above</td><td colspan="4">See above</td></tr>
<tr><td colspan="4">Excellent, if space permits</td><td colspan="4">Excellent, if space permits</td><td colspan="4">Undetermined</td><td colspan="4">Excellent, but may be wasteful of space</td><td colspan="4">Limited only by the search time and cost</td><td colspan="4">Limited only by the search time and cost</td></tr>
<tr><td colspan="4">Excellent</td><td colspan="4">None</td><td colspan="4">Undetermined</td><td colspan="4">Excellent</td><td colspan="4">Excellent</td><td colspan="4">Usually requires re-recording</td></tr>
<tr><td colspan="4">Excellent</td><td colspan="4">Excellent</td><td colspan="4">Undetermined</td><td colspan="4">Good</td><td colspan="4">Satisfactory, but may not be necessary</td><td colspan="4">See above</td></tr>
<tr><td colspan="4">Microfilming and ADP type skills</td><td colspan="4">Microfilming</td><td colspan="4">Microfilming</td><td colspan="4">Electronics and ADP type skills</td><td colspan="4">ADP operators</td><td colspan="4">ADP operators</td></tr>
<tr><td colspan="4">Good</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">INA</td><td colspan="4">Excellent</td><td colspan="4">Excellent</td></tr>
<tr><td colspan="4">Partial</td><td colspan="4">None</td><td colspan="4">Partial</td><td colspan="4">Good</td><td colspan="4">Excellent</td><td colspan="4">Excellent</td></tr>
<tr><td colspan="4">Essential</td><td colspan="4">Essential</td><td colspan="4">Essential</td><td colspan="4">Essential</td><td colspan="4">Essential</td><td colspan="4">Essential</td></tr>
<tr><td colspan="4">Excellent, if space permits</td><td colspan="4">Only if space permits</td><td colspan="4">limited to some systems only</td><td colspan="4">Only if space permits</td><td colspan="4">Limited only by the search time and cost</td><td colspan="4">Requires a special procedure</td></tr>
<tr><td colspan="4">Excellent, but time consuming</td><td colspan="4">Very limited</td><td colspan="4">See above</td><td colspan="4">Excellent</td><td colspan="4">Excellent</td><td colspan="4">Usually requires re-recording</td></tr>
<tr><td colspan="4">Satisfactory by adding new cards</td><td colspan="4">Satisfactory by adding new cards</td><td colspan="4">See above</td><td colspan="4">Excellent by adding new pages</td><td colspan="4">Limited only by the search time and cost</td><td colspan="4">Requires a special procedure</td></tr>
<tr><td colspan="4">Requires refilming</td><td colspan="4">Requires refilming</td><td colspan="4">limited to some systems only</td><td colspan="4">Good, by erasing and re-recording</td><td colspan="4">Excellent</td><td colspan="4">Usually requires re-recording</td></tr>
</table>

EDGE NOTCHED CARD	OPTICAL COINCIDENCE	EAM PUNCHED CARD ③	MISC. CARD SELECTORS	MICROFILM JACKET	MICROFICHE
DR X, DS, URS, DFR -X	DF X, DS, URS, DFR -X	DR X, DS, URS, DFR -X	DR, DS X, URS, DFR	DR, DS X, URS, DFR	DR, DS X, URS, DFR
Up to 10	Up to 25	Up to 15	Up to 200	100 up	100 up, total of all locations
None	Only if more than one set is kept	None	None	Good	Good
Good	Good	Very limited	Very limited	Good	Excellent
None	None	None	None	None	None
None	Limited	Limited	Limited	Good	Good
Limited	Limited	None	None	Excellent	Excellent
Good	Very Good	Very Good			
1 usually	As many as desired, usually	1 to 4 normally			
Extensive	Extensive	Extensive	Nominal	100%	100%
Up to 9 for each term searched	2 for each term searched	Varies, but usually extensive	2 if the document no. is known	4 or more	4 or more
Knowledge of needle sorting	None	ADP operations	None	None	None
Document nos., misc. written, graphic	Document number address	Any data recorded on cards	Same as that for record medium	Page images	Page images
Direct viewing	Visual display or printout, see D & D	EAM cards or printout listing	Same as that for record medium	Viewer screen, paper or film copy	Viewer screen, paper or film copy
			None	Full pages usually	Full pages usually
20 minutes up	2 minutes	20 minutes up	Less than 10 sec. if document no. is known	Less than 1 minute	Less than 1 minute
			5-60 sec, may need an extra unit	Less than 1 minute	Less than 1 minute
			Up to 30 sec., may need an extra unit	Less than 1 minute	Less than 1 minute
No	Yes in some systems	Yes	To manual methods only	To microfiche or aperture card only	None
Very limited	Excellent	Good	Same as that for record medium	Good, but time consuming	Excellent
Limited	Excellent	Intermediary is usually required	Good, but not always practical	Satisfactory	Excellent
Good	Excellent	None	Good, but not always practical	Satisfactory	Good
None	None	Limited			
Tiresome, if heavilly used	Very easy	Tiresome if used extensively	Excellent	Good, but subject to user resistance	Good, but subject to user resistance
Satisfactory	Excellent	Excellent	Satisfactory	Satisfactory	Satisfactory
Satisfactory	Satisfactory	Excellent	Satisfactory	Satisfactory	Satisfactory
Satisfactory, usually	Very Good	Subject to wear	Same as that for record medium	Good	Very Good
Satisfactory	Satisfactory	Excellent	INA	Satisfactory	Satisfactory

PART B - RETRIEVAL AND PRESENTATION

<table>
<tr><th colspan="2" rowspan="3">CHARACTERISTIC OR FEATURE</th><th colspan="4">MICROFILM STRIP</th><th colspan="4">MICROFILM ROLL MECHANIZED</th><th colspan="4">MICROFILM ROLL PHOTO-OPTICAL</th><th colspan="4">MICROFILM CHIP AUTOMATED</th></tr>
<tr><th>DR</th><th>DS</th><th>URS</th><th>DFR</th><th>DR</th><th>DS</th><th>URS</th><th>DFR</th><th>DR</th><th>DS</th><th>URS</th><th>DFR</th><th>DR</th><th>DS</th><th>URS</th><th>DFR</th></tr>
<tr><td></td><td>-X</td><td></td><td>-X</td><td></td><td>X</td><td></td><td>-X</td><td>X</td><td>X</td><td>X</td><td>-X</td><td></td><td>X</td><td>-X</td><td></td></tr>
<tr><td rowspan="6">ACTIVITY</td><td>IDEAL AVERAGE NO. OF RETRIEVAL ACTIONS PER DAY</td><td colspan="4">DS-under 100
DFR- 500 up</td><td colspan="4">100 up</td><td colspan="4">Varies - often inverse to info volume</td><td colspan="4">DS- 500 to 700
URS- 1000 up</td></tr>
<tr><td>CAPABILITY - SIMULTANEOUS USE BY TWO OR MORE PEOPLE</td><td colspan="4">DS- Good
DFR- None</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">Limited</td></tr>
<tr><td>CAPABILITY - LOW COST DECENTRALIZATION TO USER LOCATIONS</td><td colspan="4">Limited</td><td colspan="4">Good</td><td colspan="4">Very limited</td><td colspan="4">None</td></tr>
<tr><td>CAPABILITY - DIRECT QUERYING FROM REMOTE SITES</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">Depends on the equipment used</td></tr>
<tr><td>CAPABILITY - HANDLING LARGE PEAK LOADS</td><td colspan="4">Limited</td><td colspan="4">Limited</td><td colspan="4">Limited</td><td colspan="4">OK, if batching is permissable</td></tr>
<tr><td>PORTABILITY</td><td colspan="4">Limited</td><td colspan="4">Limited</td><td colspan="4">None</td><td colspan="4">None</td></tr>
<tr><td rowspan="5">INTELLECTUAL</td><td>CAPABILITY - CORRELATING AND MANIPULATING STORED DATA</td><td colspan="4"></td><td colspan="4"></td><td colspan="4">Good, unless multi rolls are hinderance</td><td colspan="4">DS- None; URS- same as the host computer</td></tr>
<tr><td>NO. OF DESCRIPTIVE TERMS THAT CAN BE SEARCHED AT SAME TIME</td><td colspan="4"></td><td colspan="4"></td><td colspan="4">6 to 15</td><td colspan="4">Same as above</td></tr>
<tr><td>HUMAN INTERVENTION INVOLVED IN SEARCH OR LOOK-UP PROCESS</td><td colspan="4">100%</td><td colspan="4">100%</td><td colspan="4">Nominal</td><td colspan="4">Nominal</td></tr>
<tr><td>NO. OF STEPS REQUIRED PER RETRIEVAL ACTION</td><td colspan="4">4 or more</td><td colspan="4">4 or more</td><td colspan="4">3 or more</td><td colspan="4">4 or more</td></tr>
<tr><td>SPECIAL SKILLS REQUIRED IN USING THE SYSTEM (1)</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">Machine searching</td><td colspan="4">Microfilming and machine operations</td></tr>
<tr><td rowspan="3">PHYSICAL</td><td>OUTPUT - TYPE OF INFO OR DATA FURNISHED BY THE SYSTEM</td><td colspan="4">Page images</td><td colspan="4">Page images</td><td colspan="4">Page images</td><td colspan="4">Page Images</td></tr>
<tr><td>PRESENTATION OR DISPLAY METHOD</td><td colspan="4">DS- paper copy
DFR- view screen</td><td colspan="4">View screen or a paper copy</td><td colspan="4">View screen or a paper copy</td><td colspan="4">View screen and/or paper or film copy</td></tr>
<tr><td>CAPABILITY - RETRIEVAL OF PORTIONS, ONLY, OF DOCUMENTS</td><td colspan="4">DS- whole document
DFR- full page</td><td colspan="4">Full pages, usually</td><td colspan="4">Full pages only</td><td colspan="4">Full pages only</td></tr>
<tr><td rowspan="9">SERVICE</td><td>RESPONSE TIME - FROM INITIATION OF QUERY UNTIL USER VIEWS RESULTS</td><td colspan="4">DS- 30 min. up
DFR- less than 30 sec.</td><td colspan="4">half to 2 minutes</td><td colspan="4">30 sec. plus 15 sec. per 1000 pages</td><td colspan="4">DS- under 30 sec.
URS- few seconds</td></tr>
<tr><td>RESPONSE TIME - OBTAINING PAPER COPY OF SINGLE PAGE OR SHEET</td><td colspan="4"></td><td colspan="4">Less than 30 seconds</td><td colspan="4">Less than 30 seconds</td><td colspan="4">Under 30 seconds</td></tr>
<tr><td>RESPONSE TIME - OBTAINING MICROFILM COPY</td><td colspan="4"></td><td colspan="4"></td><td colspan="4">Any copy feature, not for use of the user</td><td colspan="4">Under 10 seconds, where available</td></tr>
<tr><td>CONVERTABILITY TO OTHER METHODS AND EQUIPMENT</td><td colspan="4">None</td><td colspan="4">Microfilm jackets and strips only</td><td colspan="4">Usually jackets and film strips</td><td colspan="4">Aperature cards, in some systems</td></tr>
<tr><td>CAPABILITY - PRODUCTION OF DUPLICATE OR ALTERNATE SYSTEM</td><td colspan="4">Good</td><td colspan="4">Good</td><td colspan="4">Good</td><td colspan="4">Good</td></tr>
<tr><td>CAPABILITY - USER SELF SERVICE</td><td colspan="4">DS- very limited
DFR- limited</td><td colspan="4">Limited</td><td colspan="4">Very limited</td><td colspan="4">Very limited</td></tr>
<tr><td>CAPABILITY - DIRECT BROWSING</td><td colspan="4">Good</td><td colspan="4">Excellent</td><td colspan="4">Limited</td><td colspan="4">None</td></tr>
<tr><td>CAPABILITY - USE FOR CURRENT AWARENESS</td><td colspan="4">Limited</td><td colspan="4">None</td><td colspan="4">None</td><td colspan="4">None</td></tr>
<tr><td>PHYSICAL EASE IN USING THE SYSTEM</td><td colspan="4">DS- awkward
DFR- satisfactory</td><td colspan="4">Good, but subject to user resistance</td><td colspan="4">Good, but varies with diferent equipment</td><td colspan="4">Nominal</td></tr>
<tr><td rowspan="4">QUALITY</td><td>ACCURACY OF RESULTS (2)</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td><td colspan="4">Excellent</td><td colspan="4">Excellent</td></tr>
<tr><td>EQUIPMENT RELIABILITY</td><td colspan="4">Good</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td></tr>
<tr><td>SECURITY AND DURABILITY OF RECORDING MEDIA</td><td colspan="4">Very good</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td></tr>
<tr><td>EQUIPMENT DURABILITY</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td><td colspan="4">Satisfactory</td></tr>
</table>

MICROFILM EAM PUNCH CARD	MICROFILM NOTCHED CARD	MICROFILM SUPERMINIATURE	VIDEO TAPE	COMPUTER-ALL (3)	COMPUTER MASS MEMORY (3)
DR / DS / URS / DFR	DR / DS / URS / DFR	DR / DS / URS / DFR	DR / DS / URS / DFR	DR / DS / URS / DFR	DR / DS / URS / DFR
DS: X; URS: -X	DS: X; URS: X	DS: X	DS: X; DFR: -X	DR: X; DFR: X	DR: X; DFR: X
DS- 100 up URS- up to 15	DS- Up to 100 URS- up to 10	Undetermined	1000 up	Varies with data volume, service, etc.	Varies with data volume, service, etc.
DS- excellent URS- None	DS- excellent URS- none	Very limited	None	Possible, but not always practical	Depends on the host computer
DS- excellent URS- very limited	Excellent	Good	Limited	Very limited	Very limited
None	None	INA	Good	Possible, but not always practical	Depends on the host computer
DS- good URS- limited	DS- good URS- none	Limited	Limited	Often requires batching	Depends on the host computer
DS- good URS- none	Limited	Good	None	None	None
Very good	Good		None	Excellent	Depends on the host computer
1 to 4 usually	1 usually		None	No limit, usually	Depends on the host computer
Extensive	Extensive	Varies widely	Nominal	Nominal, except for remote query	Depends on the host computer
DS- 4 or more URS- numerous	DS- 4 or more URS- 6-9 per term	3 or more	3 or more	Varies widely	Varies widely
Microfilming and ADP operations	Knowledge of needle sorting	Microminiature tech.	Machine Operations	ADP programming	ADP programming
Page image	Page image	Page image	Page image	Computer stored or generated data	Computer stored or generated data
View screen, paper or film copy	View screen, paper or film copy	Usually view screen only	View screen and paper copy	Punch card, printout, and video	Punch card, printout, and video
Full page only	Full page only	Full page only	Full page only	Excellent	Depends on the host computer
DS- less than 1 min. URS- 20 min. up	DS- less than 1 min. URS- 20 min. up	Less than 1 min.	Varies, average is 1 minute	Varies with each system	Varies with each system
Under 30 sec.	Under 30 sec.		Varies with print equipment	Varies with each system	Varies with each system
Under 1 min.	Under 1 min.				
Yes	DS- Yes URS- No	INA	None	Possible, but may be costly	Possible, but may be costly
Excellent	Very limited	Excellent	Good	Excellent	Excellent
DS- Excellent URS- Very limited	Limited	Excellent	Excellent	Limited, at present	Depends on the host computer
DS- Good URS- None	Good	Good	Excellent	Limited, at present	Limited, at present
Limited	None		None	Excellent	
DS- may be resisted URS- can be tiring	DS- may be resisted URS- can be tedious	Good, but subject to user resistance	Good, but subject to user resistance	Very good	Very good
Excellent	Good	Satisfactory	INA	Excellent	INA
Excellent	Good	INA	INA	Usually very good	INA
Unit record machines may damage film	Satisfactory	Satisfactory	Image quality may drop off	Requires special care and supervision	Requires special care and supervision
Excellent	Excellent	INA	INA	Satisfactory	INA

PART C - RESOURCES *([illegible] completed by the person making the study)*

	FACTOR	CLUE-WORD EXTRACT CARD				PERMUTED INDEX				COLUMNAR CARD				DUAL DICTIONARY			
		DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
		X				X				X				X			
INITIAL COSTS	PERSONNEL TO DESIGN AND ASSIST IN INSTALLATION OF THE SYSTEM																
	PERSONNEL TO SUPERVISE SYSTEM AND PROVIDE REFERENCE SERVICE																
	PERSONNEL TO ENTER INFORMATION INTO THE SYSTEM																
	INFORMATION ACQUISITION COSTS																
	EQUIPMENT COSTS AND SERVICE CHARGES																
	SUPPLIES, SPACE, AND MISC. COSTS																
CURRENT CAPABILITY *(availability of)*	PERSONNEL TO DESIGN AND ASSIST IN INSTALLATION OF THE SYSTEM																
	PERSONNEL TO SUPERVISE SYSTEM AND PROVIDE REFERENCE SERVICE																
	PERSONNEL TO ENTER INFORMATION INTO THE SYSTEM																
	INFORMATION TO BE ENTERED INTO THE SYSTEM																
	EXTERNAL INFORMATION SERVICES																
	EXISTING MECHANIZED EQUIPMENT OR ADP SERVICES																
	EXISTING INFORMATION AIDS OR TOOLS																
	EXISTING COMMUNICATION AND TRANSPORTATION SERVICES																
	SUITABLE SPACE																
	ELECTRIC POWER																

EDGE NOTCHED CARD				OPTICAL COINCIDENCE				EAM PUNCHED CARD				MISC. CARD SELECTORS				MICROFILM JACKET				MICROFICHE			
DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
X			-X	X			-X	X			-X		X				X				X		

PART C - RESOURCES (continued)

	FACTOR	MICROFILM STRIP				MICROFILM ROLL MECHANIZED				MICROFILM ROLL PHOTO-OPTICAL				MICROFILM CHIP AUTOMATED			
		DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
			·X		·X		X		·X	X	X	X	·X		X	·X	
INITIAL COSTS	PERSONNEL TO DESIGN AND ASSIST IN INSTALLATION OF THE SYSTEM																
	PERSONNEL TO SUPERVISE SYSTEM AND PROVIDE REFERENCE SERVICE																
	PERSONNEL TO ENTER INFORMATION INTO THE SYSTEM																
	INFORMATION ACQUISITION COSTS																
	EQUIPMENT COSTS AND SERVICE CHARGES																
	SUPPLIES, SPACE, AND MISC. COSTS																
CURRENT CAPABILITY (availability)	PERSONNEL TO DESIGN AND ASSIST IN INSTALLATION OF THE SYSTEM																
	PERSONNEL TO SUPERVISE SYSTEM AND PROVIDE REFERENCE SERVICE																
	PERSONNEL TO ENTER INFORMATION INTO THE SYSTEM																
	INFORMATION TO BE ENTERED INTO THE SYSTEM																
	EXTERNAL INFORMATION SERVICES																
	EXISTING MECHANIZED EQUIPMENT OR ADP SERVICES																
	EXISTING INFORMATION AIDS OR TOOLS																
	EXISTING COMMUNICATION AND TRANSPORTATION SERVICES																
	SUITABLE SPACE																
	ELECTRIC POWER																

MICROFILM EAM PUNCH CARD				MICROFILM NOTCHED CARD				MICROFILM SUPERMINIATURE				VIDEO TAPE				COMPUTER-ALL				COMPUTER MASS MEMORY			
DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR	DR	DS	URS	DFR
	X	-X			X	X			X				X		-X	X			X	X			X