Not For Publication

HIGHWAY RESEARCH REPORT

DOCUMENT RETRIEVAL IN A HIGHWAY RESEARCH LIBRARY

George B. Sherman and Robert A. Anderson

Presented At The 48th Annual Meeting
of The Highway Research Board

January 1969

68-21

STATE OF CALIFORNIA

TRANSPORTATION AGENCY

DEPARTMENT OF PUBLIC WORKS

DIVISION OF HIGHWAYS

MATERIALS AND RESEARCH DEPARTMENT

RESEARCH REPORT

NO. M & R 631140

Prepared in Cooperation with the U.S. Department of Transportation, Bureau of Public Roads April, 1968

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

State of California

Transportation Agency
Department of Public Roads
Division of Highways
Materials and Research Department

DOCUMENT RETRIEVAL IN A HIGHWAY RESEARCH LIBRARY

by

George B. Sherman
Assistant Materials and Research Engineer

Robert A. Anderson
Librarian, Materials and Research Dept.

Presented at the 48th
Annual Meeting
of the Highway Research Board

January 1969

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

ABSTRACT

REFERENCE: Sherman, G. B., and Anderson, R. A., "Document Retrieval in a Highway Research Library", State of California Department of Public Works, Division of Highways, Materials and Research Department. Research Report No. 631140, April, 1968.

ABSTRACT: A document storage and retrieval system in the California Division of Highways, Materials and Research Department Library is described. The system was designed and installed as part of a research project whose objectives were to evaluate a concept-coordinated document retrieval and storage system for internally and externally produced documents, utilizing the work of the Highway Research Information Service of the Highway Research Board. The HRIS list of authorized terms has been adopted for indexing, and an optical coincidence retrieval method has been installed. Input, storage, searching and output operations are described. It is concluded that this system is feasible and economical in this application.

KEYWORDS: Libraries, information systems, information retrieval, documentation, indexes/locators/, index terms, documents, highways, highway research, California. Division of Highways. Materials and Research Department.

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

TABLE OF CONTENTS

	Page
INTRODUCTION	1
INFORMATION NEEDS IN THE DEPARTMENT	3
FACTORS INVOLVED IN ADOPTING THIS SYSTEM	5
BASIC EQUIPMENT	7
INPUT OPERATIONS	8
SEARCH OPERATIONS	9

LIST OF FIGURES

	Figure
BASIC EQUIPMENT OF THE SYSTEM	1
INPUT OPERATION	2
KEYWORD CARD FILE	3
ALPHABETICALLY FILED KEYWORD LOCATOR FILE	4
DOCUMENT WORK SHEET	5
PLACING KEYWORD CARDS ON LIGHT SOURCE	6
KEYWORD CARDS SUPERIMPOSED ON LIGHT SOURCE	7
AN EXAMPLE OF A SUPERIMPOSABLE CARD SYSTEM SEARCH	8

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

INTRODUCTION

The task of recovering information from written records of research work in the field of highway transportation has increased enormously in the past several years. For example, the Highway Research Board has increased its annual published output of technical papers from about 1,000 pages to about 10,000 pages in the past 20 years. This increase has resulted from the expansion of research and development activities in the highway transportation field.

The published literature in almost any field shows a similar growth. In 1905 the American Concrete Institute published 114 pages of technical information. In 1966, a total of 3,309 pages was reached. Retrieval of this information on demand is essential not only for research personnel carrying out their daily work but for management decisions which must be made.

The Highway Research Board's computer-based Highway Research Information Service (HRIS) is now in operation. There are over 14,000 references stored on computer tapes, and over 10,000 new items are expected to be added each year. The HRIS has in its publication, HIGHWAY RESEARCH IN PROGRESS, covered ongoing research in the United States and, through cooperative efforts, research activities in the rest of the world. The recent publication, HRIS ABSTRACTS, also provides abstracts of published articles from a wide range of the periodical and proceedings literature.

On a local level, however, a similar situation exists in the retrieval of information. Usually operated by a small staff, the specialized highway library is finding it a more difficult task to cope with the flood of material received. The specialized highway library is finding it more difficult to index and retrieve information from the technical literature using traditional methods, especially from a subject viewpoint. For example, indexing a document using the traditional methods of subject headings with a typed card for each subject access point is an imposing problem in card reproduction.

As a case in point, it is decided that a document provides information on eight subjects. If this document were processed in the traditional manner, ten typed cards would be added to the card index to represent author, title, and eight subject access points. To provide ten access points for 1,000 documents, 10,000 cards would be added to the card index, and, theoretically, many of these cards would have to be consulted to find documents containing the desired information.

ClibPDF - www.fastio.com

With a view toward utilizing some of the newer methods of document retrieval the California Division of Highways, in a research project, proposed to try the newer methods. Its objectives were to evaluate the feasibility of using a concept-coordinated storage and retrieval system for research information storage and retrieval for the Materials and Research Department Library and other libraries of the Division of Highways. The scope of the report is limited to experience in the California Division of Highways, Materials and Research Department Library.

The authors feel that this research project has proven in great measure a saving in both man-hours and money to the Division of Highways.

From the standpoint of the engineering staff the simplicity and ease of "following up" and pinpointing information is an on-the-spot operation.

The rewarding factor in the system is that it gives the seeker of certain information a wider range of usable or applicable data. In a sense, it has acted as a catalytic agent in following a trend of thought to a successful search. In other words, by the accessibility of records covering a many faceted subject we are gaining in immeasurable ways. Needless to say, there is nothing more disconcerting to the operator of an information system than to receive an inquiry so nebulous that a good deal of time is consumed in attempts to extract possible leads pertaining to his subject. The "mental gymnastics" incurred in questioning this information system are simplified and the answers to questions posed are quickly made available.

The availability of the HRIS Thesaurus list of indexing terms was a great factor in adopting a coordinate indexing system since it allows the questioner to pose an inquiry in natural language. There is no need for a technical question to be converted into computer language in order to use the system.

The opinions, findings, and conclusions expressed in this publication are those of the authors and not necessarily those of the BUREAU OF PUBLIC ROADS.

ClibPDF - www.fastio.com

INFORMATION NEEDS IN THE DEPARTMENT

The requirement for information in the Materials and Research Department can only be understood in the total environment. Users are individuals with real requirements, facing real problems, requiring real solutions. "A's" needs for information may be vastly different from "B's" needs and at a later date, and in response to an entirely different situation. An examination of these differing requirements may be in order.

1. SPECIFIC DOCUMENT.

   The most frequent request received is for a specific document usually requested by title, author, or number. The requester knows exactly what he wants and any traditionally organized document collection would satisfy his needs.

2. SEARCH FOR A SPECIFIC FACT.

   Searches are made in this case for a particular fact or item of information--a formula, an equation or details of when something occurred. The user does not wish to consult all references that contain this fact. One source will suffice. An automated retrieval system is not the most efficient way of locating such specific information.

3. RETROSPECTIVE SEARCH.

   When a research project is initiated, or a paper or lecture is being prepared, the requester must review the literature relating to his subject in order to avoid repeating the work of others. The user usually asks for all pertinent references in the system. It is here that a system built on new techniques of filing and retrieving documents containing relevant information is best employed.

The traditional library filing and retrieval systems involve several catalog cards for each document which is classified by subject in the Dewey Decimal, Universal Decimal, Library of Congress, or other specialized classification system. The documents are then filed in accordance with their subject classification. Catalog cards are filed alphabetically, using authors and subject headings as filing points.

ClibPDF - www.fastio.com

The vast increase in non-traditional forms of publishing such as preprints, technical reports, in-house research project reports and laboratory reports has made the traditional library methods of multiple catalog cards and classification systems cumbersome and expensive methods of information retrieval for a small information center.

The index to a document collection is the prime component in information retrieval systems. The Materials and Research Department's index to stored information is carried on optical coincidence cards. Some criteria for selection of this type of equipment are:

1. User oriented - the index is controlled and manipulated by the user.

2. Rapid retrieval - the index is in close proximity to the file and actual retrieval of the document is rapid.

3. Simplicity - the selection of keyword cards and superimposition of these cards is a simple task. Addition or subtraction of keywords is easily done.

4. Open ended - the size of the collection is of no importance. Addition of documents is in numerical order.

ClibPDF - www.fastio.com

FACTORS INVOLVED IN ADOPTING THIS SYSTEM

In the decision to adopt this particular Information and Retrieval system the following points were considered:

1. The documents to be included

2. The terminology to be used to describe the intellectual content of the documents

3. The searching device that would be most effective and economical.

The following will explain how each of these points were decided upon:

1. The Materials and Research Library regularly circulates to the research staff technical journals, abstract bulletins, publication notices and other announcement media. This material contains notices of published information which are of current or potential value to the Department. Many documents are received automatically through distribution by other states and research agencies. Others are obtained through request, purchase or exchange. The readers of these written communications, as specialists, are in the best position to ascertain their value to the document collection. Therefore, documents are included in the information and retrieval system only if their significant values may be defined in terms of the immediate and future objectives of the Materials and Research Department. This "quality control" of documents for inclusion in the collection is just as important to the information center as quality control of materials that go into the construction of highways. The readers of these written communications are also in a unique position to suggest keywords for indexing purposes.

The preliminary screening given to incoming documents determines which staff member is best qualified to evaluate the content. Each staff member receives the reports most closely related to his specialty. This procedure features the added advantage of keeping specialists cognizant of the latest developments in their fields.

Another benefit of reading technical information with "quality control" in mind is that in their own writings, engineers and technical personnel become cognizant of the value of significant titles and meaningful abstracts.

ClibPDF - www.fastio.com

2. After the source documents have been acquired and the decision is made to include them in the system, a means of expressing the intellectual content of a document was selected. In a coordinate indexing system such as this, terminology control is provided by the HRIS THESAURUS LIST OF INDEXING TERMS. This list of descriptive terms for indexing documents in the highway field includes broad generic concepts, as well as specific concepts, to describe document content. The Thesaurus list enables the reader or indexer to assign uniform terms to represent information in the document for retrieval by a concept coordinated search.

3. In a manual system such as this a means must be selected to store the coordinate index. In a conventional system this index is comprised of the author and subject cards which direct the searcher to the desired document. The Materials and Research Department, after investigation of other methods, chose optical coincidence cards. These are punchable or drillable cards, each of which represents a keyword. The 10,000 position grid imprinted on the card provides the storage of document numbers or positions in the file. Reports and documents are indexed by keywords relating to the information concept of the contents. Cards identified by these keywords are precision drilled. Each hole designates a particular document.

To retrieve documents on specific subjects, the card or cards having the proper keywords are placed over a light source and allow spots of light to shine through coincident holes, which then are read on the imprinted grid. The system offers pinpoint selectivity of technical concepts, from which it takes its name as a concept retrieval system.

The searcher can determine very quickly the result of a too broad or constrained search by the number of light spots appearing. The search can then be broadened or narrowed by adding or subtracting keyword cards

In the following paragraphs the equipment and process of entering a document into the system and search procedure will be discussed.

ClibPDF - www.fastio.com

BASIC EQUIPMENT

The basic equipment of this system is shown in Figure 1. The encoder-viewer consists of a light box and frame on which is mounted an electric drill attached to a sliding carriage which provides movement along both vertical and horizontal axes. There is a 0-99 scale on each axis, with stops at each position allowing the drill to be set precisely at any one of 10,000 grid positions. Lowering the drill and pressing the foot switch allows us to drill a hole in one or more keyword cards in a position representing the accession number of a specific document. Figure 2 shows the drilling operation. Guides on the machine facilitate cards to be precision placed. The drilling surface is made of a translucent plastic illuminated from beneath. This aids in the drilling operation and may be used for read-out purposes. To enable the input and search operations to go on simultaneously, there is an auxiliary viewer which consists of a light source behind a translucent screen and provided with guides for aligning the keyword cards. This auxiliary viewer is shown in Figure 1.

Shown in Figure 3 is the keyword card file. Each 11-$\frac{1}{2}$" x 10-$\frac{1}{4}$" card has a 10,000 position grid imprinted thereon. The file consists of ten decks of 100 cards each. The file dividers are placed at 100 card intervals. Through the use of distinctive colors and numbered tabs, a total of 1,000 color positions enable a random access. To locate and refile keyword cards it is just necessary to file in the correct deck of 100 cards and the correct color.

Since the keyword cards are filed randomly, they are located by alphabetically filed cards. These cards are contained in an ordinary rotary-type file. (See Figure 4) These locator file cards contain the keywords which have been used for the document collection and the number of the keyword card. The photograph shows that the keyword NUCLEAR MOISTURE-DENSITY DETERMINATIONS is No. A-02 in the keycard file.

When a search is conducted, this rotary file is used to locate the correct keyword card(s).

ClibPDF - www.fastio.com

INPUT OPERATIONS

Reports to be entered into the system are given an accession or file number in consecutive order from 0 to 9999. This is a non-significant number except for filing and retrieval purposes. It is not a classification number which reflects the content of the document.

With each document is a document work sheet. (Figure 5.) This work sheet is patterned after that of the Department of Defense. (After more experience this work sheet may be modified.) The document and the work sheet bear the same number.

The corporate and personal author of the document are indicated on the work sheet. The typist then types conventional "3 x 5" catalog cards to be filed into the library card catalog. These cards serve to answer inquiries when either the corporate or personal author are known.

The indexer then assigns keywords which reflect the subject content of the report and indicates the keycard number. The keywords are selected from the HRIS Thesaurus. However, articles in increasing number are arriving in the library already indexed with keywords and these are used when applicable.

The worksheet is then taken to the input device, the keycards for the document being entered are removed from the keycard file and a hole drilled through all of the cards in the grid position dedicated for that document. The keycards for that document are then refiled into the keycard file. In practice about 25 documents are entered each time the input process takes place.

It is to be noted that the document work sheets becomes an integral part of the storage system. These work sheets are filed in binders in accession number order. When a search is made the work sheet may be consulted instead of the document. This may satisfy the request or direct the questioner to another file or another record location. This completes the input process.

When 10,000 documents are entered into the system a new set of keyword cards is started for the second 10,000 documents.

ClibPDF - www.fastio.com

SEARCH OPERATIONS

Exploitation of this system is based on a search technique which requires that a number of general classes be combined to produce a more specific class. Thus, the search procedure starts with a "translation" of the search question into terms of the input vocabulary. The HRIS Thesaurus contains guides for this translation, such as: PUMPCRETE USE PUMPED CONCRETE. The vocabulary also contains the reciprocal reference PUMPED CONCRETE USED FOR PUMPCRETE. In this way the search is directed to the term PUMPED CONCRETE for information and away from PUMPCRETE.

As an actual query to the system, let us say the inquirer asks if among the documents in the file there is a bibliography of articles on black bases. Turning to the Thesaurus under BLACK BASES the searcher finds the notation BLACK BASE - USE BITUMINOUS AGGREGATE BASE. There is also a term BIBLIOGRAPHIES. Combining the two term cards BITUMINOUS AGGREGATE BASE and BIBLIOGRAPHIES over the auxiliary viewer it is apparent immediately that in the file document number 1763, a bibliography on Bituminous Aggregate Bases, should satisfy the questioner. The coincidence of holes viewed in front of a light source indicates documents which satisfy the question. (Figure 6 and 7)

The Thesaurus not only serves as the input vocabulary, but also as a means of translating the inquiry into index terms used by the system.

Another example illustrating the principles of a coordinated search is shown in Figure 8. The questioner asks, "What research projects done by the Materials and Research Department led to the development of TEST METHOD CALIF. 229, "Method of Test for Durability Index". Referring to Figure 8, the keyword cards AGGREGATES, DURABILITY, TEST METHODS, MATERIALS AND RESEARCH DEPARTMENT are removed from the keyword card file and superimposed on a light source and numbers 13 and 27 show a light spot. These documents should contain the desired information.

In the case of a retrospective search the request might be "I want all the information in the file on the corrosion of reinforced concrete bridges in a marine environment". The proper keyword cards are superimposed and the numbers of the relevant documents are read and then removed from the file.

The advantage of the optical coincidence system is that the whole file may be searched and pertinent document numbers read out very rapidly.

ClibPDF - www.fastio.com

If too many light spots appear it is easily possible to place one or more keyword cards on the auxiliary viewer to make the question more specific. Conversely, if none or too few light spots appear one or more cards may be removed to generalize the question. This feature of being able to "talk" to the system has proven to be of great use.

To summarize, it appears that this system has proven to be effective in enabling the Department to index, store and retrieve both internally and externally produced documents at a low cost for equipment. It has proved an effective tool to use in conjunction with the more complete retrieval service of the Highway Research Board Information Service.

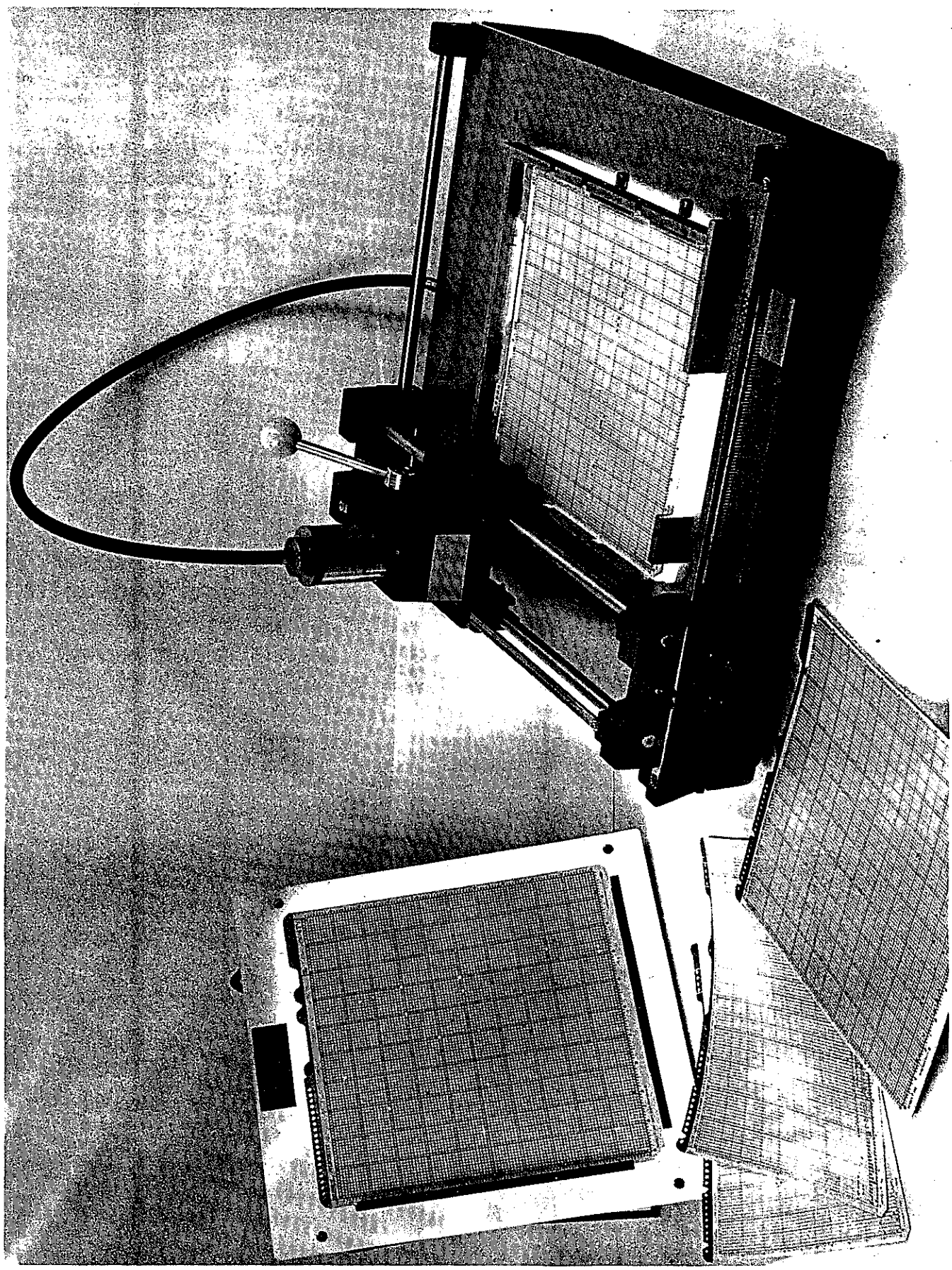
We were aided immeasurably by having a standardized vocabulary in the existence of the HRIS Thesaurus of Indexing Terms. By using this already existing vocabulary and choosing equipment which cost less than $1,500 we have a document retrieval method which affords many more and precise approaches to stored literature than the use of traditional library methods.

ClibPDF - www.fastio.com

FIGURES

ClibPDF - www.fastio.com

Figure 1



Basic Equipment Of The System

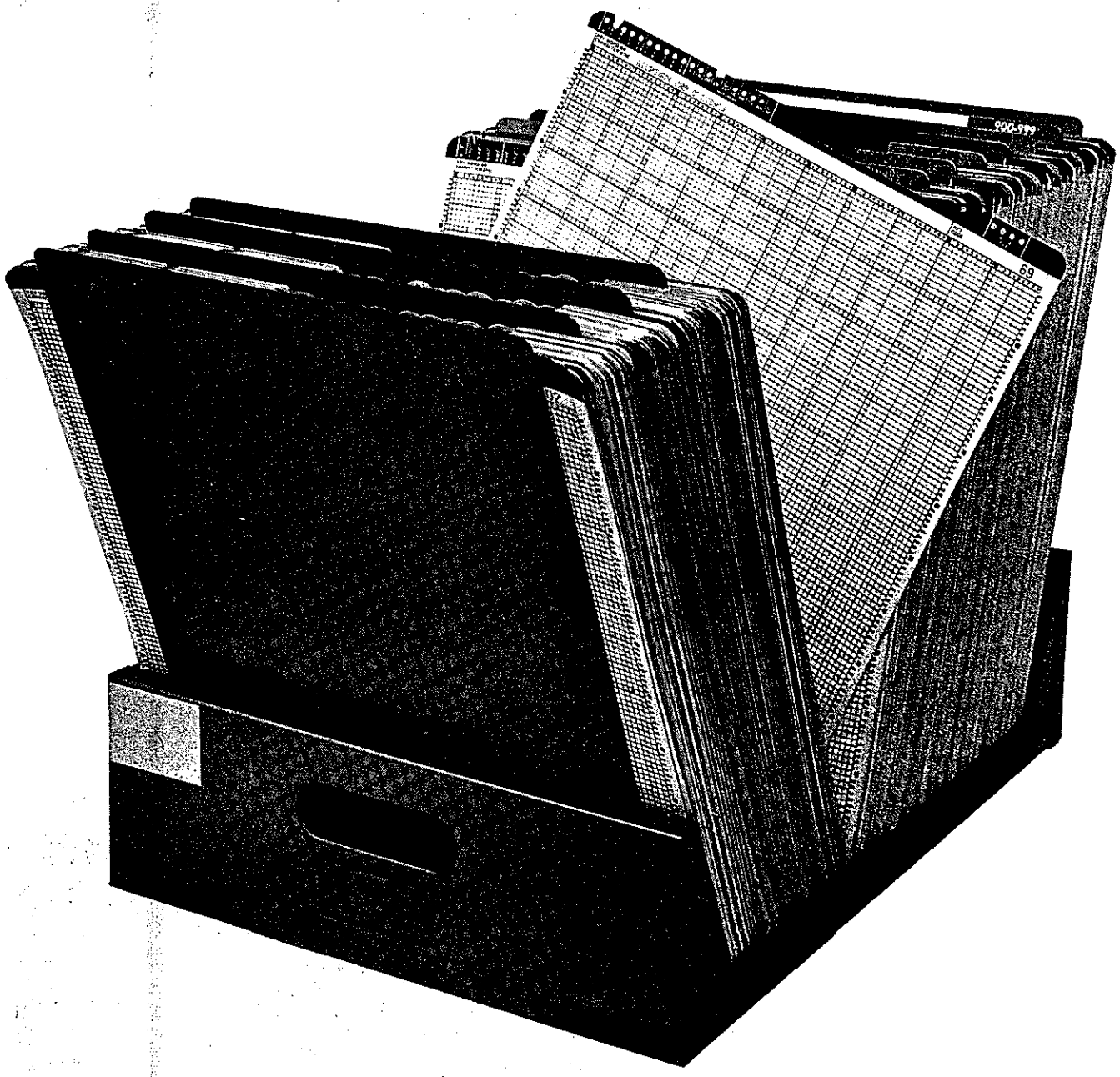
ClibPDF - www.fastio.com

Figure 2

Input Operation

ClibPDF - www.fastio.com

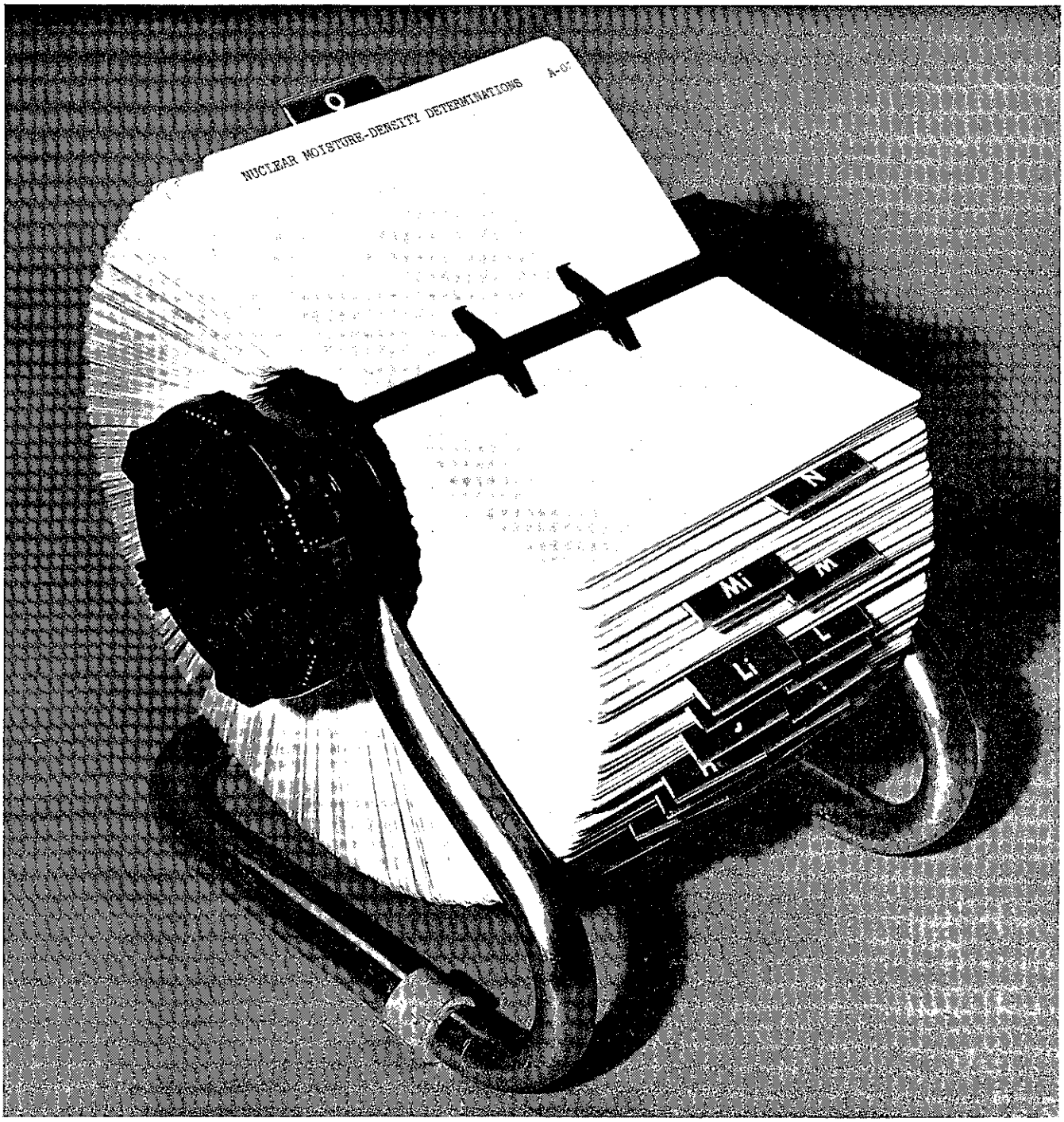
Figure 3



Keyword Card File

ClibPDF - www.fastio.com

Figure 4



Alphabetically Filed Keyword Locator File

ClibPDF - www.fastio.com

Date 2-27-68 Figure 5 Accession No. 0407

DOCUMENT WORK SHEET

1. ORIGINATING ACTIVITY *(Corporate Author)*

California. Division of Highways. Materials & Research Dept.

2. REPORT TITLE:

Use of Nuclear Soil Gages in Construction Control

3. DESCRIPTIVE NOTES *(Type of Report and Final Dates)*

Research Project

4. AUTHOR(S) *(Last Name, First Name, Initial)*

Smith, T.W.; Weber, W.G.; Howe, D.R.

5. REPORT DATE	6. TOTAL NO. OF PAGES	7. NO. OF REFS.
1964-66		

8. ORIGINATORS REPORT NUMBER(S)	9. OTHER REPORT NO(S) *(Any other numbers that may be assigned this report)*
M & R Research File No. 2602	

10. AVAILABILITY M & R LIBRARY YES ☐ NO ☒ WHERE

M & R Research File No. 2602

11. SUPPLEMENTARY NOTES	12. SPONSOR

13. ABSTRACT/SUMMARY

Phase I - A backscatter type nuclear surface moisture-density gage was successfully employed as the sole means of compaction control on a large earth fill project in District 01.

Phase II - A similar gage was used on several projects at one time in Districts 08 & 10 in order to determine if one gage could be used simultaneously on more than one project for compaction control.

HMR T-1148 (REV. 2/68)

Document Work Sheet

ClibPDF - www.fastio.com

Figure 5 (continued)

14. KEY WORDS

Construction Control A-104

Compaction A-00

Compaction Control A-01

Nuclear Moisture-Density Determinations A-02

Embankment Compaction A-03

Soil Compaction A-04

California Division of Highways, Materials & Research Dept. A-05

Nuclear Applications A-861

15. REPORTS ISSUED

Howe, D.R.; Weber, W.G.

1. The Application of a Nuclear Soil Gage to Construction Control (District 01) - January 1966

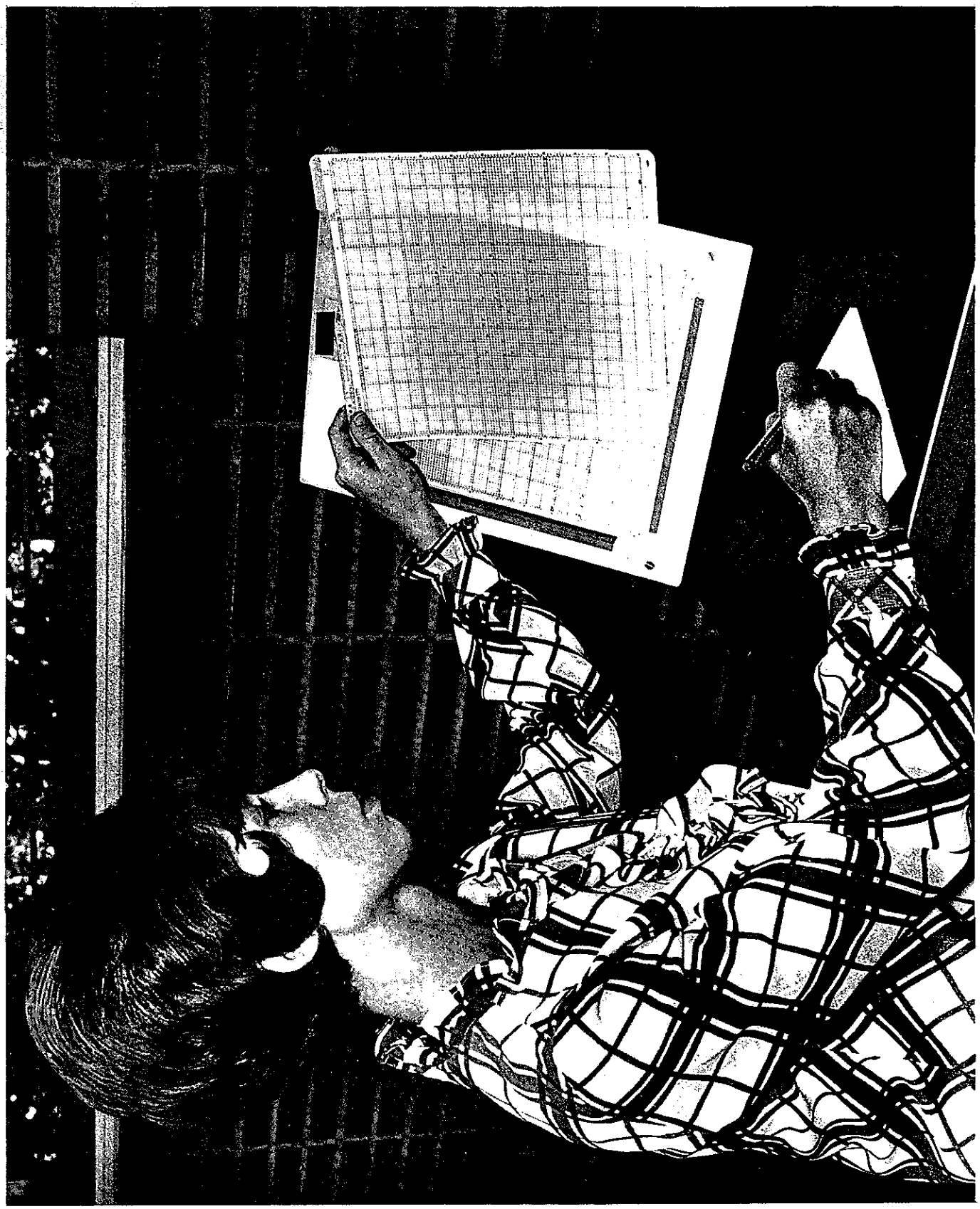
2. Report on the Field Use of a Nuclear Soil Gage on Several Concurrent Construction Projects in Districts 03 and 10 August 1965

Smith, T.W.; Howe, D.R.

Control of Embankment Compaction by Nuclear Testing - Jan. 1966 Printed Eighteenth Annual Calif. Street and Highway Conference of the ITTE University of California - Berkeley - January 27-29 1966

ClibPDF - www.fastio.com

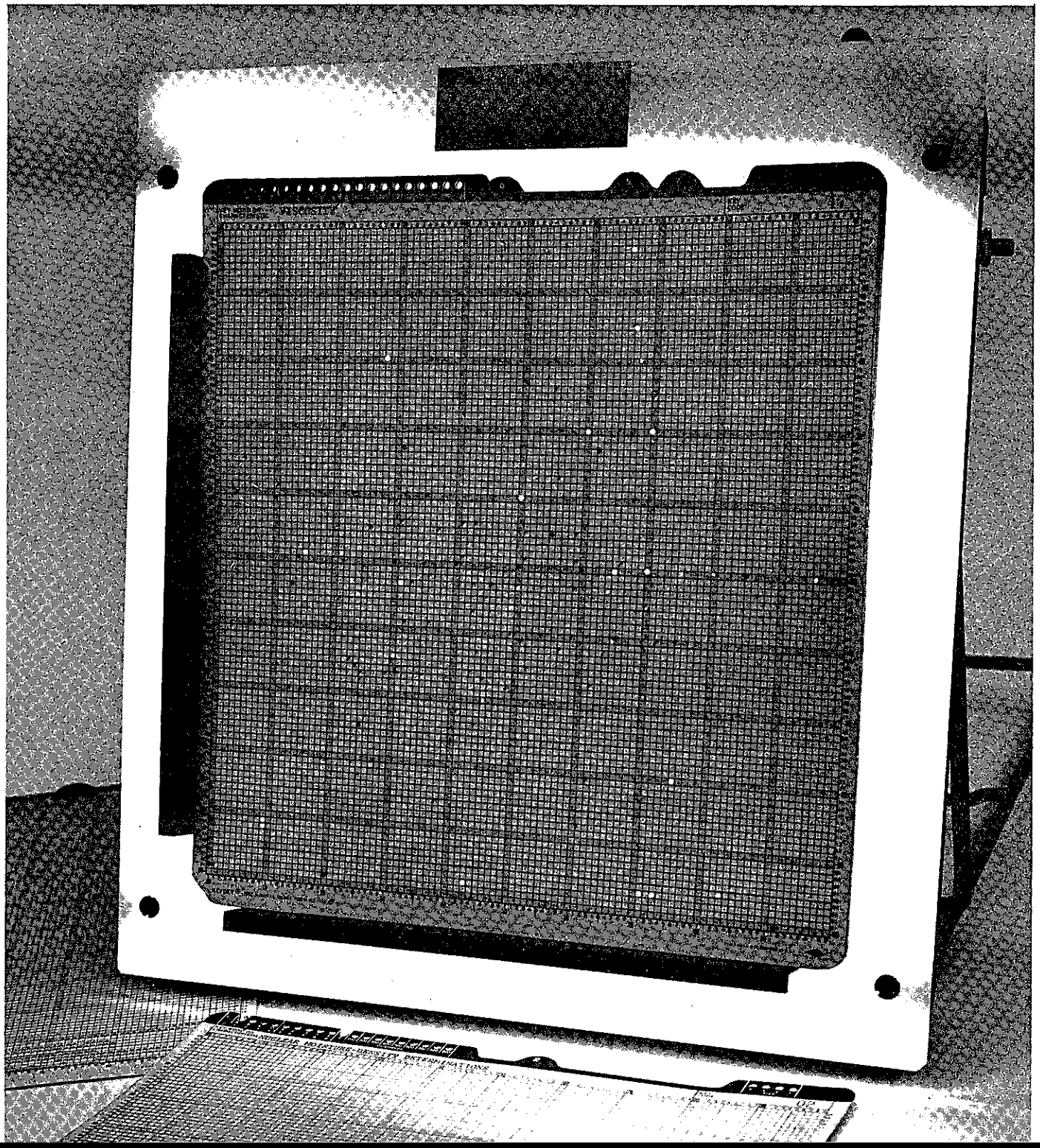
Figure 6



Placing Keyword Cards On Light Source

ClibPDF - www.fastio.com

Figure 7



ClibPDF - www.fastio.com

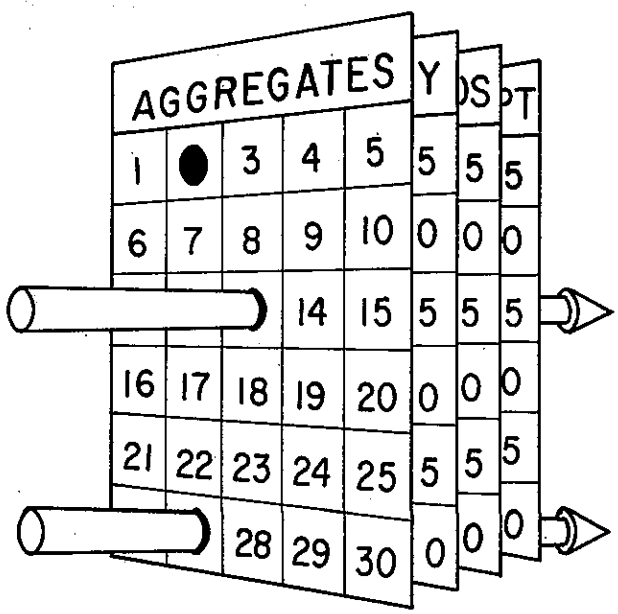
Figure 8

AGGREGATES				
1	●	3	4	5
6	7	8	9	10
11	12	●	14	15
16	17	18	19	20
21	22	23	24	25
26	●	28	29	30

DURABILITY				
1	2	3	4	●
6	7	8	9	10
11	12	●	14	15
16	17	18	19	20
21	22	23	24	25
26	●	28	29	●

TEST METHODS				
1	2	3	4	5
6	●	8	9	10
11	12	●	14	15
16	17	18	19	20
21	22	23	24	25
26	●	28	●	30

MAT'L.& RES. DEPT.				
1	2	3	4	5
6	7	8	●	10
11	12	●	14	15
16	17	18	19	20
21	22	23	24	25
26	●	28	29	30



An example of a superimposable card system search. Each item of information has a position corresponding to its location in a serially ordered file. A search is conducted by superimposing a number of "term cards" and observing coincidence of holes. These documents should contain information on the desired subject.

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com

ClibPDF - www.fastio.com