UC
LECTURES

SPRING
1963

University of Califomia--Berkeley
School of Librarianship

October 1963

Course Outline
L245 (Intro. to Documentation)
Chas. P. Bourne, Instructor

I Course Content and Scope, Procedures,

What is documentation and documentalists?

Conventional Libraries have no monopoly on File Problems

(Engineering drawings, Patents, FBI, mailing lists, business files)

World's Technical Lit. Problem (Quantitative and Qualitative--effect

of Big Science)

New and Improved Tools--Topics for the rest of the course

II Classification and Indexing--The Organization of Information (several sessions)

Definitions (doc. retrieval vs. ref. retr. vs. info. retr.;

file; file item)

Care and Feeding of Large Files--Some Fundamental Problems

Organization of the File

Continuum of Methods for Subject Analysis

Specific Methods of Subject Analysis

Methods of Subject Index Display

III Coding--The Indexing Shorthand

General

Statistical Nature of English Words

Methods for Abbreviating English Words and Names--Derived Coding

Assigned Coding

Prime Number Coding

Superimposed Coding

IV Machine-Language Representation

General

Tab Cards

Paper Tape

Magnetic Tape

MICR

OCR

V Manual Card Systems

General

Edge-Notched Cards

Interior-Notched Cards

VI Tab Card Systems

Equipment

Applications

Costs

VII Methodology of Information System Design

General

Problem Definition and Determination of User Requirements

Synthesis and Design

Evaluation

VIII Computers

Equipment

Programming

Applications

General

Analysis of Text Structure (Concordances, Thesauri,

Text Schematics)

Preparation of Retrieval Tools (indexing, abstracting)

File Searching

SDI

Copy Editing

Costs

IX Paper Tape and Magnetic Media Equipment

Paper Tape (Applications, Cost, etc.)

Magnetic Tape

X Microforms and Mechanized Image Systems

Background

Film Viewing

Hard Copy Outputs

Basic Microforms

Mechanized Image Systems

Applications

Spring 1963

MEETING NO. 1

The Nature of the Problem

INSTRUCTOR'S BACKGROUND--We can learn from each other

COURSE CONTEXT--Methods of Information Handling (Devices, Techniques, Methodology). Punched card and computer techniques, methods of information system design and analysis, special indexing tools.

CLASS PROCEDURES--Informal lectures, considerable class participation, demonstrations and practice in equipment use, project work. No text yet.

DEMONSTRATION of PCMI card, Bio-Chem Activities, and computer search printout--These all have something in common that is basic to the whole point of this class: technology is going to affect the way in which the libraries operate and perform their tasks. This course will give an awareness of this technology, as well as tools and techniques for evaluating and critiquing it. This will help all of you--regardless of what type of library you go to work for.

WHAT ARE DOCUMENTALISTS? (Lunatic fringe of the library field?) Definitions (documentation, info. specialists, etc.) Concerned w/ documents? w/ spec. libraries? Started by people outside formal library field.

CONVENTIONAL LIBRARIES HAVE NO MONOPOLY ON FILE PROBLEMS. Other organizations now have the same kinds of problems that were first faced by libraries (see fig.) Tools developed by early libraries can help with other special file problems.

Examples of file problems: (Show a commercial or special problem to correspond to a traditional library one)

Berkeley Pub. Lib., U.C., Harvard, L.C. (traditional)

Medical Records (Palo Alto, V.A.)

Eng. Drawings (80,000/plane)

Patents

CIA

Mailing lists (Time-Life)

FBI & other police files (270,000 arrests/yr. by L.A.P.D., FBI gets 25,000 record requests/week)

DOD aerial photo

L.A. Title Insurance

Soc. Security

DOD-DAC

WORLD'S TECHNICAL LITERATURE PROBLEM: Forms of lit.(reports, papers, conferences, patents, books), system of primary and secondary publications and report literature, volume and nature of the traffic. Used to be able to keep up with the papers, and then the abstracts--now there are even too many abstracts (e.g. Bio-Chem Section of Chem Abstracts have 50,000 abstracts/yr.) and there is some movement to micro-abstracts (e.g. Cardio-Vascular Index)

NEED NEW AND IMPROVED TOOLS. This course aims to help you to develop methods, tools, and systems. Very little emphasis to be given on how to operate an info center.

WHAT THE REST OF THIS IS ABOUT. Outline

ASSIGNMENT FOR NEXT SESSION - Reading below.

SUGGESTED READING FOR NEXT MEETING

1. G. Hardin, "The Last Canute," Scientific Monthly pp. 203-208 (Aug. 1946)
2. H. P. Luhn, "Key-Word-in-Context Index for Technical Literature (KWIC Index)," American Documentation, Vol. 11, No. 4, pp. 288-295 (October 1960).
3. R. A. Kennedy, "Library Applications of Permutation Indexing," J. Chem. Doc. Vol. 2, No. 3, pp. 181-185 (July 1962).
4. J. C. Costello, Jr., "Uniterm Indexing Principles, Problems, and Solutions," American Doc. Vol. 12, No. 1, pp. 20-26 (January 1961).

Spring 1963

MEETING NO. 1

The Nature of the Problem

INSTRUCTOR'S BACKGROUND--We can learn from each other

COURSE CONTEXT--Methods of Information Handling (Devices, Techniques, Methodology). Punched card and computer techniques, methods of information system design and analysis, special indexing tools.

CLASS PROCEDURES--Informal lectures, considerable class participation, demonstrations and practice in equipment use, project work. No text yet.

DEMONSTRATION of PCMI card, Bio-Chem Activities, and computer search printout--These all have something in common that is basic to the whole point of this class: technology is going to affect the way in which the libraries operate and perform their tasks. This course will give an awareness of this technology, as well as tools and techniques for evaluating and critiquing it. This will help all of you--regardless of what type of library you go to work for.

WHAT ARE DOCUMENTALISTS? (Lunatic fringe of the library field?) Definitions (documentation, info. specialists, etc.) Concerned w/ documents? w/ spec. libraries? Started by people outside/not library field.

CONVENTIONAL LIBRARIES HAVE NO MONOPOLY ON FILE PROBLEMS. Other organizations now have the same kinds of problems that were first faced by libraries (see fig.) Tools developed by early libraries can help with other special file problems.

Examples of file problems: (Show a commercial or special problem to correspond to each traditional library one)

Berkeley Pub. Lib., U.C., Harvard, L.C. (traditional)

Medical Records (Palo Alto, V.A.)

Eng. Drawings (80,000/plane)

Patents

CIA

Mailing lists (Time-Life)

FBI & other police files (270,000 arrests/yr. by L.A.P.D., FBI gets 25,000 record requests/week)

DOD aerial photo

L.A. Title Insurance

Soc. Security

DOD-DAC

WORLD'S TECHNICAL LITERATURE PROBLEM: Forms of lit.(reports, papers, conferences, patents, books), system of primary and secondary publications and report literature, volume and nature of the traffic. Used to be able to keep up with the papers, and then the abstracts--now there are even too many abstracts (e.g. Bio-Chem Section of Chem Abstracts have 50,000 abstracts/yr.) and there is some movement to micro-abstracts (e.g. Cardio-Vascular Index)

NEED NEW AND IMPROVED TOOLS. This course aims to help you to develop methods, tools, and systems. Very little emphasis to be given on how to operate an info center.

WHAT THE REST OF THIS IS ABOUT. Outline

ASSIGNMENT FOR NEXT SESSION - Reading below.

SUGGESTED READING FOR NEXT MEETING

1. G. Hardin, "The Last Canute," Scientific Monthly pp. 203-208 (Aug. 1946)
2. H. P. Luhn, "Key-Word-in-Context Index for Technical Literature (KWIC Index)," American Documentation, Vol. 11, No. 4, pp. 288-295 (October 1960).
3. R. A. Kennedy, "Library Applications of Permutation Indexing," J. Chem. Doc. Vol. 2, No. 3, pp. 181-185 (July 1962).
4. J. C. Costello, Jr., "Uniterm Indexing Principles, Problems, and Solutions," American Doc. Vol. 12, No. 1, pp. 20-26 (January 1961).

MEETING NO. 2

CLASSIFICATION & INDEXING--THE ORGANIZATION OF INFORMATION

DEFINITIONS: Document Retrieval vs. Reference Retrieval vs. Information Retrieval

File Item (book, patent, policy, etc.) = smallest module of information to be handled as a unit.

File = homogeneous collection of a single type of file item.

THE CARE & FEEDING OF LARGE FILES--Some Fundamental Problems

Certain fundamental problems occur in varying degree in nearly all files

1. Variances in indexing or cataloging (between people, by the same person), Why? reflecting different backgrounds of indexers, degree of slack in the indexing system, relative importance of the subject content, jargon and special terminology of the material.

How to Improve? Take slack out of the indexing system (authority lists, etc.), index each item by several different persons with different subject backgrounds to get composite indexing, on-the-job training.

2. Variances in File Searching--Same as above.

3. Indexing Discontinuities Caused by Personnel Changes

Why? File organization often influenced by background, interests, and procedures of the person responsible for it.

How to Improve? A good written description of the system and procedures should be provided. For small files, it may be easier to let newcomer re-organize into a "comfortable" setup.

4. Loss of Material--More difficult to recover as file gets larger Some files require absolute integrity, and steps taken to try to provide this (e.g., catalog card rod, roll microfilm). Unit-record systems generally more susceptible.

5. Difficulty of Modifying Original File Structures--Any file suffers growing pains and must change its structure as it grows--this is difficult (intellectually and physically). Intellectual problems: correct the structure for oversights, and plan for changes in subject fields, terminology, and unknown requirements. Physical problems: how to re-mark and re-position the file items, how to change the indexes and catalogs (without interrupting service).

Some large files have a full time staff member to work on this problem (collecting data on file and index traffic, etc.)

6. Loss of Familiarity with File Contents--Indexer may lose familiarity with contents and structure of a file--especially if he does not use it.

How to improve? Let the indexer use the file (e.g., routine searches, historical or state-of-the-art reviews using the file material). This is especially good for novice indexers.

7. Loss of Familarity with Subject Matter--Indexing quality for complex subjects (e.g. chemistry, electronics, etc.) usually governed by indexer's subject knowledge. Indexer tends to grow stale if he remains an observer rather than participant (less aware of new technology and needs of users).

How to improve? Rotate personnel between information and technical work, use regular technical staff on part-time basis, assign file responsibility (but not operation) to technical specialists.

8. File ~~Arranging~~ Purging--Need a program of continuous and systematic review rather than infrequent assaults. (In some business firms these purges often remove tons of paper. Statute of limitations govern the purging of business files--file usage may govern library purging practice. Info "half-lifes." No such thing as out-of-print books.

9. Snag Files--How would you get along without "misc."? Operate and treated just like splinter political parties and exist for the same reason.

ORGANIZATION OF THE FILE

One of the first considerations--manual or mechanized--and crucial to both. Mechanized Systems can operate with any method of file organization, and need not dictate how the file shall be organized, and vice versa. After choosing method of file organization, further questions are those of notation or coding and display--these are covered later. File designer has 3 degrees of freedom: method of subject analysis, notation, and method of display. These can be treated almost independently of each other. (Some systems have become known primarily by their notation, e.g., Dewey Dec., UDC, Colon.)

A CONTINUUM OF METHODS FOR SUBJECT ANALYSIS

Is one indexing system better than another? Why? In What Way? What are the fundamental differences between indexing systems? Key parameter seems to be the degree of control of the descriptor language --and the "power" of an indexing system seems to be directly related to this. So does the amount of necessary intellectual effort, and difficulty of mechanization. Extremely difficult to determine precise difference between subject headings, Descriptors, Uniterms, and other indicators of subject content.

Word-indexing vs. Concept Indexing, and all the steps in between.

SPECIFIC METHODS OF SUBJECT ANALYSIS

1. Words lifted from title or text (simplest) a) Indicate subject content by list of words extracted from the text. Only control is to provide a list of junk words. Search tool is a display showing which documents used each words. This was early basis for Uniterm system ("no vocabulary control necessary--just let it grow"). Soon begin to get problems of terminology (synonyms, generic relationships, syntactic relationships, etc.) and required improved vocabulary control. Scores of libraries have adopted this system, and it forms the basis for much of the machine indexing.

b) Description of basic Uniterm system. (No subordinate terms, deep indexing, no standard authority list, vocabulary grows with the system, accession numbers).

c) ASTIA 7000 Uniterms correspond closely with the 13,000 AEC subject headings

d) KWIC indexes utilize this same principle--but use larger junk lists (distinguishing feature is its display. Context gives slightly more "power" than plain basic system.) Description of KWIC indexes.

e) These are first draft indexes. Utility depends on how well the words describe the contents. Some indexers add words to the title before machine processing.

f) KWIC bulk--only 10-25 titles per page (average)

MEETING NO. 3

CLASSIFICATION AND INDEXING--THE ORGANIZATION OF INFORMATION (cont.)

SPECIFIC METHODS OF SUBJECT ANALYSIS(cont.)

1. Summary of Previous Points
   a) methods of file org., notation, and display, are essentially separate problems, and can be treated independently of each other.
   b) there is a continuum of methods for subject analysis (word indexing vs. concept indexing) with varying degrees of control of the descriptor language.
   c) "Power" of an indexing or classification seems to be related to the degree of control of the descriptor language, so does the amount of required intellectual effort, and difficulty of mechanizing or automating the operation.
   d) method of mechanized implementation of an indexing system is essentially independent of the indexing system chosen. The equipment needn't dictate what system is to be used, and vice versa.
2. Description of the first mechanized KWIC system (poorest language control of any present system)
   a) the only control is a list of words that are not to be used--rather than a list of allowable terms.
   b) thus any word in the universe (except 300-1000 junk words) could be used for indexing.
   c) usually operate only on titles--which are often poor indicators of content.
   d) some editors have been adding words to the KWIC title to get a better index, while asking future authors to write better titles.
   e) KWIC really amounts to a bulky first draft index--think of it in this light, and you won't oversell it.
   f) Now used for many organizations as title announcement publications, rather than retrospective search tools,--especially where the titles are particularly descriptive and meaningful (e.g. Dissertations in Physics).

3. Description of the Early Uniterm Systems

a) Rules: Determine key words which represent the subject content of the item being indexed. These words needn't appear in the text.

Record the key words on the serial accessions card so that every retrieval word is a filing word in the posting operation..

Terms are added to the vocabulary as they occur.

Where true synonyms occur, enter the well-known form and make a see reference from the other form.

b) Description of typical manual system--input and searching

c) Relatively deep indexing (i.e. large no. of access points).

d) the terms are independent of each other, and no term is subordinate to any other term.

e) Relatively easy to do complex logical searches (AND, OR)

f) Typical vocabularies of 4000-7000 Uniterms.

g) "Rose by any other name"--Uniterm may be essentially the same as a subject heading. ASTIA's present 7000 uniterms correspond closely with the 13,000 subject headings used in the AEC, and correspond closely to the subject headings in the particular Library of Congress file that ~~has~~ its ~~readers~~--thus a new name doesn't necessarily make it a fundamentally different system.

4. Description of the Clue Word System

a) As with KWIC, only key words from the text are normally used for indexing.

b) Differs from KWIC primarily in method and amount of display.

c) Display consists of a unit-record called an EXTRACT, which is an abbreviation or condensation of the original document, made by lifting key sections of the original. Indexer (using pencils as the only tools) marks the original document. The marked sections are then transcribed to another unit-record form (e.g. 5 x 8 card).

d) Copies of the new card are made and filed under each key word marked (names, organizations, products, etc.).

e) Copies of a card may be filed under ~~org.~~ 10-15 locations for the items marked.

f) Each card contains enough text to be useful, subject heading is shown in the context in which it is used. Search results normally consist of copies of these cards.

into a form useful for the developer. But this may not be the same as the viewpoint of another user, and the system, may be relatively poor for that user.

b) e.g. St. Bernard dog found in separate classifications for the animal kingdom, burglar alarms, beasts of burden, liquor transport system, or an infinity of different classifications.

c) No single way to organize entire universe of information into an ordered and well-structured pattern of relationships that is best for all users. Most methods of file organization represent a single approach or point of view. (And this should be the point of view and method of organization that suits most of the users.)

d) It is possible to analyze the file item from several points of view, and synthesize a composite classification scheme. This is often called faceted classification.

e) With faceted schemes, indexing is done separately and independently for separate schemes, and then combined together.

f) Typical facets might be property, object, method.

g) Ranganathan's Colon Class. provides a methodology and method of notation to display index terms together in a coordinate relationship. Selection by a single aspect is possible. Different facets for different subject fields.

h) WRU Semantic Code is similar to Colon Class.--differences primary in notation. Colon designed for visual reference, Semantic code designed for machine implementation. Semantic factors represents a general conceptual area of the item being encoded. Not used anywhere but WRU.

i) Faceted class. of Class-Research Group in England.

j) UDC has a facet device and point of view indicators to add to Main UDC number to suggest a particular point of view of the item (e.g. theoretical, practical, economic and commercial, use, management, etc.)

METHODS OF SUBJECT INDEX DISPLAY

1. Many different ways to display the same degree of indexing--this provides different operating features or characteristics, but cannot change the retrieval performance or power.

2. Many "different" indexing systems are actually variations only in the display or method of presentation.

3. Unit-Record Displays

   L.C. catalog card (sold 42 million in 1962, used 1½ mill. at L.C.)

   Other forms of catalog card (e.g. side-margin card)

   Clue Word card

4. Unit-Term Displays

   Uniterm cards

5. Book-Form or Published Displays

   Book catalog (IBM and Rochester)

   KWIC (could add see and see also references)

6. Planned Schemes

   Alphabetic Subject Headings (e.g. Appl. Mech. Rev.)

   Classified Subject Headings (e.g. J. Acoust. Soc. of Amer.)

   Correlative Index (e.g. Chem Abstracts)

   Modifier Lines (e.g. NSA)

   Scan-Column Index

   Tabledex

7. Other Special Indexes

   Citation Index (e.g. Shepard's, Garfield, Tukey, AIP)

   Current Contents

8. Criteria for Selection of an Indexing System

   a) No best system for all jobs

   b) Things to consider: type of user, characteristics of the collection, availability of other indexes, complexity and required accuracy of searches to be conducted, number of searches to be conducted, required response time, current user and operator attitudes toward present system, available resources for developing and maintaining the system.

Meeting No. 5

CODING: THE INDEXING SHORTHAND

GENERAL

1. System designer must decide how to ~~request~~ represent information--on paper, or in machine-readable form.
2. Coding = shorthand representation of the indexing information, and helps reduce clerical effort and storage requirements.
3. Choice and design of coding is especially important for machine systems.
4. Will discuss several general schemes now, and others later in context of a particular system.

STATISTICAL NATURE OF ENGLISH WORDS

1. Characteristic frequency distributions exist for different aspects of different types of words. This information can be exploited in the system design.
   a. Initial letters of text (T, A, O); subject words (S, P, C, T); names (B, S, M, H, C).
   b. Terminal letters.
   c. Composite distributions; (text ETOANI).
   d. bigrams (text: TH, ER, ON, AN).

METHODS FOR ABBREVIATING ENGLISH WORDS AND NAMES--DERIVED CODING

1. Many instances where this could be useful (addresses, journal or book titles, names, etc.) to save storage and communication facilities. Example: ACS Codens, KWIC ELCO notes.
2. To be applied automatically, the method must be systematic, and should have certain characteristics (minimum storage, max. discrimination power).
3. Coding possibilities with alpha letters (26^3 = 17,576, $26^5 \approx$ 12 mill.).
4. Special computer programs written to generate 3- and 4-letter words.
5. Methods: Truncation, Elimination of Vowels, N^{th} Letter Dropout, Dropout by Frequency Distribution, Dropout by Bigram Frequency Distribution, Elimination of Redundancies.
6. Use of ~~Chords~~ Check Digit:

   A B L A T I O N 1 2 9 15 14 = 59 = 26 + 26 + 7 = H

   A B L H

   This method would give a different 4-character abbreviation for ablating.

ASSIGNED CODING

1. Requires the use of detailed code books for coding and decoding.
2. Any material can be efficiently coded this way (words, sentences, paragraphs)--e.g. Bloomer's Commercial Cryptograph (1824) for cost savings and information security:

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

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

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

3. A similar Code: Adams Cable Codex (1913)

Loranthus: Sell out at once; do not delay.

Reversible: We all unite in wishing you a very Merry Christmas.

4. Commercial, government, and ham radio and teletype operators use similar code terms--for the same reasons.

GM: Good Morning

73: Best Regards

BCNU: I'll be seeing you.

5. Military organizations have standardized abbreviations for memos and note taking.

M Tk Bn: medium tank battalion

6. Some data processing systems use special name codes to identify customers and accounts. These codes may be made up of fragments of the customer's name, address, shipping route, and billing information. Use Life example (40 million names on their mailing list).

7. One and one-quarter million different names in the United States, but 50% of the people are described by 1,500 names (Smith, Jones, etc.)

8. Library example of assigned coding is the notation used with many classification schemes--for convenience in transcription and recording. For example

Dewey	621.312 136	Wind-driven electric plants
LC	TK 1001	General transmission of power
UDC	551.507.362.2	Artificial Satellites

9. Could let variable-length alphabetic data be ~~requested~~ represented by a fixed-length numeric code. This facilitates machine handling.

   Examples: ASTIA Descriptor Codes, Internal Revenue Service.

10. Use of empirically derived table to code names or descriptors into equal interval groups. Useful for deciding how to break a general population into equal-interval groups before the group is encountered.

    Example: Partition of UC student body into equal-interval groups for registration hours.

    e.g. A - Breec
         Breed - Culp
         Culq - Fram
    } each will account for about 10% of the people encountered

    Example: Cutter-Sanborn Tables (1880) are a partition of author names into equal-interval groups in order to establish a range of call numbers that are uniformly spread across a range of book authors. Other equal-interval tables, e.g. Sayer's Manual of Class for Librarians and Bibliographers (1955) have also been devised for library code numbers.

PRIME NUMBER CODING

1. Definition of prime number. For coordinate indexing schemes, let each descriptor be represented by a prime number, and the composite index statement by the product of all pertinent prime numbers. Subject index would then consist of a list of numbers (one per file item). Searches are made by dividing search primes

into composites, and looking for file numbers which result in no remainder.

e.g. Automatic indexing = 13

Bliss Classification = 23

Cataloging = 29

A document indexed by these three terms would have the index notation of 13 x 23 x 29 = 8671. There is no other way to get this number except by multiplying these three numbers together. Consequently, division by any number other than these three will result in some remainder. Only these three can be divided into this composite without getting a remainder. The division and test for remainder is the search ~~question~~ operation.

2. Difficult and awkward for manual use, but suitable, and perhaps convenient for machine searching.

3. May get some very large numbers if the descriptor vocabulary and/or depth of indexing is large.

e.g.

50th prime =	227
100th prime =	523
1,000th prime =	7,907
5,000th prime =	48,593
10,000th prime =	104,723

With a dictionary of 10,000 terms, a file item with 12 terms would have a number slightly less than $(104,723)^{12}$ which is ~~requested~~ represented by about 61 decimal digits. A smaller average number could be achieved by assigning lower primes to the most common terms.

4. This procedure has been used with a few computer searching systems.

SUPERIMPOSED CODING

1. Dfn: Using binary (number base with digits 0 and 1 only) code patterns for each descriptor, as specified by a code table, superimpose the patterns for the file items relevant descriptors, in order to obtain a single composite code pattern. For example:

magnetic	0100001
digital	1000010
computer	1010000
circuit	0000011
Composite index notation:	1110011

The resulting numbers describes the indexing of the file item.

2. Code patterns are usually assigned with uniform random distribution to distribute the marking uniformly over the code field.

3. Search is done by testing for pattern inclusion.

4. Possibility of false drop. (Example) This is minimized by proper code design (longer field, etc.).

5. Primarily used to date, for edge-notched card systems, but can also be used with IBM cards, computers, and other systems.

6. With a given number of marks per descriptor, the number of marks in the composite descriptor will depend on the number of descriptors superimposed. Distributions and bounds can be estimated.

Meeting No. 6

MACHINE-LANGUAGE REPRESENTATION

GENERAL

1. For machine retrieval or processing, some numeric, alphabetic, or special symbols must eventually be transformed to some physical representation--such as holes in a card or magnetized spots on a magnetic tape--in order that they can be stored and interpreted by the equipment.
2. Most equipment has a limited capability for recognizing or handling such symbolic information.
3. Most methods were worked out independently by various manufacturers--introducing problems of compatability (e.g. IBM and Remington Rand card codes).

TAB CARDS

1. IBM (predominently) and some Remington Rand for essentially all U.S. systems. Other types in Europe.
2. Cards come in a variety of sizes (51- , 60- , 66- , 80- column for IBM; 32-, 40-, 54-, 66-, and 90-column for Remington Rand) but most people use the largest size possible.
3. Available in variety of color codes, and are usually pre-printed for specific tasks, "Scored" cards and card sets (e.g. retail slips) are also available.
4. Problems of card handling (folding, stapling, humidity effects, etc.)

5. Symbols entered onto card by:

   a. Manual key punching

   b. Automatic punching

   c. Mark Sensing (un-hurried point of origin recording for small amount of data)--27 character/card

   d. Pushing out pre-scored positions (Post-A-Punch)--40 char./card

6. Card Codes (80 x 12 = 960 for IBM; 540 for Remington Rand) have lots of potential capability, but restricted by manufacturers. Most equipment ~~estimated~~ restricted to limited character set.

   a. Direct Coding (most common)--one-for-one correspondence between source data and punched data (e.g. names, house numbers, telephone directories)

   b. Abbreviated Coding (for efficient use of card ~~codes~~ columns)--condense basic information (e.g. abreviate state names)

   c. Category Coding (more efficient card use)--require a table look-up operation. e.g.

Subject Field	Code
Physics	01
Chemistry	02
⋮	⋮
Zoology	99

Geographic Area	Code
San Francisco	01
Los Angeles	02
⋮	⋮
New York	99

Uniterm	Code
Aerodynamic heating	1334
Diffraction	2123
Methyl Groups	3024

d. Category Coding with Unconventional Punching (~~involved~~ invalid machine codes for most equipment)--2^{12} = 4096 different patterns possible in each card col. A single col. could represent a large class of objects. e.g.

Animal	Rows punched in Column 1
Horses	3 and 4
cows	3 and 5
sheep	1 and 6
⋮	⋮
cats	7 and 8
⋮	⋮

Could even show a ~~hak~~ heirarchial structure with a single card column. e.g.

Subject	Rows punched in Column 1
Airplanes (all types)	1
" - jets	1 and 2
" - turboprop	1 and 3
" - propeller	1 and 4
Ground Vehicles (all types)	2
" " - 4 axles	2 and 3
" " - 6 axles	2 and 4
" " - 8 axles	2 and 5
Boats (all types)	3
⋮	⋮

e. Superimposed Codes (invalid machine codes for most equipment)

7. Speeds: up to 2,000 cards/minute.

PAPER TAPE

1. Physical Form (5-, 6-, 7-, 8-channel tape; edge-punched card; special tapes such as Monotype and Robotype; roll vs fan-fold).
2. Method of Punch Representation (~~chord~~ chad, chadless, electrofax)
3. Code representation - Definition of "channel" and "frame." No true standard exists although some are used more frequently than others. Patterns represent data as well as control signals (CR, ignore, ~~space~~ ~~man~~ space, etc.)
4. Speeds: 1 - 1,000 ft/sec

MAGNETIC TAPE

1. Digital vs analog recording (definitions - analogy to paper tape and strip charts)
2. Special problems with magnetic tape (susceptible to stray magnetic fields, dust, shock, heat, humidity, etc.)--Patterns are not visible to human eye and must be sensed electrically (describe basic mechanics of moving tape past the read-write heads). Tapes can be reproduced easily and accurately.
3. Physical form: rolls, cards (Mylar or plastic with sandwich coat) Rolls of 2,400 to 3,600 feet of 1/2, 3/4, or 1-inch width.
4. Code representation: Similar to paper tape (5 to 20 tracks; 100 to 1,000 pulses per linear inch of tape). No standards. Tapes written on one manufacturer's machine may not have

meaning to another manufacturer's machine.

5. Tapes can be moved at high speeds (e.g. 75-200 in./sec) for transfer rate of 25,000 - 200,000 characters/second.

MICR

1. Over 90% of U.S. banks are now encoding checks with numbers pre-printed in magnetic inks.
2. Only reading numerals plus a few special symbols--in an industry standard font.
3. Method now being extended to other applications (e.g. turn-around documents for billing, airline tickets, coupons) but has little immediate utility for library application.
4. Primarily useful for paper that gets handled by the public (smudges, over-markings, etc.)

OCR

1. Optical equivalent of mark-sensing (e.g. N.Y. Stock Exchange)
2. Several models of equipment in operation to read numeric data (e.g. turn-around documents from computer printouts)
3. Some alphanumeric readers are in operation (line and page readers)
4. Very limited capability so far (font restrictions, problems of ~~graphic~~ graphic handling, various paper sizes, etc.)

Meeting No. 7

MANUAL CARD SYSTEMS

GENERAL

1. Manual card files (e.g. 3 x 5) have been in use for years, and have been a convenient and flexible means for recording sorting, merging, and up-dating, in files where each item or unit-record consists of a relatively small amount of information and can be handled independently of all the other file items.

2. Some mechanical modifications have been incorporated to speed up or simplify the card handling (methods of notching or tagging the card edges.)

EDGE-NOTCHED CARDS

1. Definition and description of how they work. Many shapes and forms, singles or carbon sets, pre-printed for specific tasks, pilot holes or blanks, microfilm inserts, coupon books, duplicating stencils.
2. Sequence Sort vs. Selection (Description)
3. Card dimensions and number of holes is determined by the card supplier. User develops the coding and notation system for his specific task. Could even use IBM card punched around the edges.
4. Coding (use illustrations)

   (a) Numeric (by one-for-one, 7-4-2-1, SF coding to show how digit was used, triangular coding, etc.) 7-4-2-1 most efficient, but there is ambiguity. (zero vs. no-data, "1" punched for several digits)--However it does allow the deck to be needled into sequence with 4 passes. Addition of "0" and SF (or V) do not help sorting, but do help selection. Effect of SF field can be obtained by deep-punching some digits. Number ranges can be quantized.

   (b) Alphabetic (by one-for-one, number of alphabet letter position, triangular format, modifier codes, etc.)

(f) Dates and Times

(d) Names Usually select a part of a name: which part?

(e) Category Coding--In common use, to achieve economy in punching. (e.g. single hole for each state or print type)

(f) Superimposed Coding

5. Equipment

(a) McBee (Keypunch, Groover, Data Punch, Selector Frame)

(b) Zatacard Selector

(c) Electrofile

6. Applications

(a) Reporting and Charging Systems

Time cards (e.g. cannery)

Charge tickets (e.g. nursing station ticket)

Library charge-Out System

Requisition Forms

(b) Technical Data Storage

Zeco Transistor Index

Test data (e.g. maintenance records)

Material characteristics (e.g. chem data)

Parts specifications

(c) Scheduling Systems

Classroom scheduling

Production scheduling

(d) Personnel Records

Skills and Background

Report Card

(e) Bibliographies (considerable interest in this, primarily for small collections of information)

(f) Business Records

mailing lists

sales analysis

purchase orders, etc.

7. Costs: 20 times IBM card. (Pay $15,00/thous. for tab size cardss).

INTERIOR-NOTCHED CARDS

1. Needled Cards (e.g. Petroleum Research Corporation)--similar in operation to edge-notched cards

2. Optical (Peek-a-Boo) Cards

(a) Describe implementation and method of operation. Limitations because of limited number of hole positions.

(b) Representative systems (NBS, Jonkers, Rem Rand, Omnidex, Port-A-Punch, Minimatrex)

(c) Extra features (e.g. NBS Microcite, automatic equipment for punching or reading.)

(d) Rapid input and search speeds.

3. Costs: approx $2000 for machine system

Meeting No. 8

TAB CARD SYSTEMS

EQUIPMENT

1. Keypunch--describe operation for 026 and manual punches

   (duplicate, skip, control card, key in only the variable data, print data at top of column)

   2 cards/min max speed

   2% errors in cards

2. Verifier--Same keyboarding as Keypunch--describe operation.

   Reduce errors to .02%, usually with different operator.

   Machine verification vs. sight verification.

3. Reproducing Punch--For duplicating a check, or punching fixed information into another deck.

   Regeneration for worm-out leaks

   Additional decks for distribution or multiple filing or parallel processing

   Machine automatically checks old pattern with new punched pattern

   Can copy the data into other fields (i.e., 80 cols. of original card can be scattered about the new card)

   Introduce concept of plugboard. (Then can be stored intact)

   One model reads mark-sensed cards

   One model imprints the edge of the card

   One model has arithmetic unit to accumulate totals from cards, and punch this data into the same or other cards

Often used as computer output device

100-300 cards/min.

4. Interpreter--For imprinting selected information on top of card to correspond to information punched in the card

Selective printing (i.e., only chosen cols. in chosen places)

Better quality

Useful when cards are generated by equipment other than printing keypunch (e.g., as with reproducing punch)

Permits the cards to be used as file documents in some cases (e.g., aperture card systems, student class cards)

60 cards/min.

5. Sorter--for sorting or selection, usually on single-column basis.

Examine one column per card pass, and direct card to one of several hoppers, depending on value punched in that position

Numeric selection or sorting on 3 cols., requiring 3 passes

Alphabetic handling doubles the number of passes

Some sorters have mechanical counters on each hopper to provide a tally

450, 650, 1000, 2000 cards/min.

6. Collator--some logic built into this machine, it can:

select specific cards from a deck by range or inequality tests, and sequence tests

Merge 2 ordered decks into a single ordered deck

Notch one deck against another

Check the sequence of a deck.

Often used for file maintenance to merge new information into an existing file (easier than re-sorting)

1300 cards/min.

7. Line Printer (407 Accounting Machine)--prints a line at a time, instead of sequential characters

First of these machines designed as accounting machines to read information from cards, perform some computations, and print the results along with other information from the cards or plugboard wiring--now little use is made of arithmetic capability; used primarily for card listing.

Plugboard can format the card cols. onto the page from single or sequential cards--Usually print one card per line.

One model punches part of a card (24 cols.) and prints on the remainder (useful for utility billing, etc.)

Paper tape forms control (skip, slew, line spacing), and variety of paper forms (checks, bond, multilith masters, catalog cards, etc.)

120 char./line

Restricted font (upper case)

150 lines/min. (sometimes used as computer printers)

8. Tape-to-Card Punch (and vice-versa)--Description of operation.

9. Statistical Sorter--Much more flexible than single-column sorter.

Plugboard and internal logic permit complex searches to be made on a single card pass.

Can search or sort fields for ranges of variables (e.g., 0-20, 21-40, 41-80) in a single pass of the deck

Check sequence of a deck

Sort by length of field

Select every nth card

450 cards/min.

10. COMAC--operate by merging term decks to form new deck with intermediate products or sums, or complements, which is then played against the third term card, and so on.

Usually faster than sorting, since only a few cards have to be passed through the equipment.

Each xxxx card normally holds 12 6-digit document numbers.

11. Card-Activated Cameras--Many indexes, catalogs, and other publications prepared by this method. (e.g., Index Medicus, NSA, Dun & Bradstreet)

Type onto white cards with any type font or arrangement (not restricted to 80 char.).

Read the cards and photograph the top line of the cards (sometimes 2 or 3 lines).

Cards can also be punched and sorted or handled by conventional tab methods.

Photo neg. is then used to prepare master repro. copy

Useful for re-arrangement of unit-records into several different lists for publication

Useful for maintaining a catalog or other periodic publication

Cards usually typed on electric typewriters or varitypers 120-240 cards/min.

Cost 6-20 cents/line, compared with tab listing at 2 cents/line.

APPLICATIONS

1. Composition and Printing are Major Objectives

telephone directories

list of periodical holdings (e.g., SRI directory, New Serial Titles) ~~printed~~

printed book catalogs (e.g., Univ. of Rochester, L. A. County Library System)

Union lists of holdings (merging card decks from several libraries)

Catalog cards

Machine posting for Uniterm cards

KWIC Indexes (Ohlman)

2. File Searching is Major Objective

Serial scan most common method (with single-column sorter)

~~Term~~ Term cards and collators also used

Superimposed ~~edge~~ codes can be searched with statistical sorter

Selected cards can be listed to prepare bibliography.

3. Other Library Activities

Book charging and circulation control

Acquisitions

COSTS

1. Cards (1.50 per thousand; Port-a-Punch $3.00 per thousand; card sets $4-10 per thousand)

2.

Equipment	~~Punches~~ Purchase Cost	Monthly Rental	Hourly Rental
Keypunch	$3800	$60	4
Sorter	2500-10,000	40-250	6
Collator	11,000-14,000	250	7
Statistical Sorter	24,000	500	14
407	30,000-56,000	400-900	16
Reproducing Punch	6000	70-140	6
Interpreter	8000-15,000	130-300	6

Lease (80%) vs. purchase. Also maintenance cost.

Meeting No. 10

THE METHODOLOGY OF INFORMATION SYSTEM DESIGN

GENERAL

1. Very little recorded information on how a new information system is designed. (Analogous to writing on how to ride a bicycle)
2. No written rules, procedures, handbooks
3. Intent of the discussion to print out some of the useful tools and methodologies. (useful as idealized goal)
4. Dfn: Information system = Complex of people, equipment, and procedures--working together to provide needed information to a group of users.
5. Generally follow 3-step procedure:

   Prob. Dfn. and Determin. of User Requirements

   Synthesis and Design

   Evaluation

   with major and minor iterations, and shortcuts (e.g., salesman's design) to achieve re-design, re-definition of goals, re-evaluation.
6. Design cycle can be done on idea phase (i.e., paper study) or physical plane (i.e., equipment development or pilot installations).

PROB. DFN. AND DETERMINATION OF USER REQUIREMENTS

(4 different and complementary approaches. User ≡ ultimate user, not operator)

A. Statistical Studies of Quantitative Parameters

   1. Obtain quantitative measures where possible. Examples:

      File Size and Growth Rate

      No. of Service Requests

No. and Type of user to be served

No. and Type of operating personnel

Total Cost and Per Unit Cost of Providing Services

Response Times to Queries

Distribution of Age of File Items

Physical Form of Search Products Provided

2. Nothing elegant about this approach--a simple arithmetic process that can be helped sometimes by sampling techniques (e.g., reference counting to find out what journals are used).

B. Functional Analysis of Job Processes (description of processes being performed)

1. Flow charting of operations, and material or information flow through the system are useful. (Some work has been done to flow chart the library operations)

2. Detailed descriptions of worker's activities--to see if there is duplication or mis-application of effort.

3. Record equipment and facility utilization (machine load chart)

4. Forms analysis chart can illustrate processes being performed.

5. I. E. techniques useful here.

C. Identification of Relevant External Influences

1. Identify design constraints and factors over which designer has no control (e.g.: 1) No. and type of personnel available for operation; 2) time and money available for design, installation, and operation; 3) requirement to be compatible with other systems; 4) requirement to use some equipment or services that already exist in other parts of the organization.

2. Important to identify these things early in the design stage.

D. Determination of Specific User Requirements

1. Requirements that Stem from External Sources

a) Some information requirements stem from sources other than the ultimate user--may even be unknown or recognized by the user--and have nothing to do with the personal interests or wants of the users. (e.g., 1) practicing physicians and Papp test; 2) Research Managers and new programs; 3) School Director's Curriculum change and effects on teachers)

2. Identification of User Habits, Preferences, and Idiosyncracies

a) These things would make the system more comfortable to the user--but are not crucial to the means of the operation. (e.g., 1) users' preferred terminology; 2) users' prior knowledge of reference tools; and 3) various habits or gimmicks that make life easier for user).

b) These special preferences need not be absolute design goals, but they should be identified so that they may be included in the design where possible.

3. Identification of Specific Requirements (main design goals and system objectives)

a) Specific and stated requirements (e.g., 1) tolerable search response time; 2) tolerable amount of irrelevant material provided; 3) preferred intervals for current awareness reporting; 4) time span for retrospective searches; 5) preferred form of search product)

b) Many of these studies classed as opinion surveys--difficult to say what the optimal or ideal requirements might be.

c) Many methods available for determining user requirements.

Traffic Observation

Direct observation of human (material traffic can indicate what sources are used, how often, what times, and what places. (e.g.; 1) plots of withdrawal rates for U.C. Reserve Library can show when the demands will occur, and what the ranges of expected traffic might be; 2) studies of circulation records and reference lists (citation counts) might (?) indicate what sources are of most interest (value to the users.)

Direct Questioning of User

Most common approach is to directly ask questions of a user (interview or questionnaire)--easiest and most obvious--but full of inaccuracies and must be done with skill to get meaningful data. (Need good interview and questionnaire design).

Example: During the last time that you performed a search, or had a search done, how long was it from the time you made your report until you received the major group of relevant references? Was this adequate, or did you really need the material sooner? (If needed sooner, how much sooner?)

Difficulty: separate need from habit

Indirect Questioning of the User

Probe the circumstances surrounding the user's last need for information.

(re-construct the situation)

How many libraries have suggestion boxes?

Question People Who have Provided the Service

Problems: Operators can only speak for a part of the potential population (the part that comes in to use the service); biased response and inertia; lack of objectivity.

Some useful information collected by these people (e.g., search requests) can be analyzed, e.g., form, viewpoint, complexity of search question.

Problem: Danger of studying query rather than original motivating requirement (queries may be translation of request into a form that has a chance of successful response by the present system)

Diary Keeping

Record incidents of needs and use of information (monologue rather than dialogue)--comprehensive or sampled--from weeks to months.

Unstructured data may be difficult to analyze at end of test period.

Difficult to get full reporting.

The instrumentation disturbs the process being measured.

Postulate New System and Invite Requests

Allows user to ask for services not provided by present system--thus removes some constraints on his requests.

Difficult for user to project himself into this pretended environment (e.g., ask a person in 1900's what features and performance he would like to have in an automobile)

AIP has used this approach to obtain search questions useful for aid in design of indexing systems.

Concrete version--operate pilot system with new features, and observe reactions and comments.

Pose System Specs for Debate

Debate proposed features to bring out features otherwise unnoticed (useful at Univ. of Ill.)

Monitor Genesis of Information Needs of Users

Monitor group from beginning to end of a project--noting when and where information needs arise--and what the circumstances are at that time.

Controlled Experiment

Several groups on common problems with different information resources. See how their productivity and effectiveness is affected by type and amount of services provided.--Sensitivity analysis to show relative importance of various services.

Perturb Present System

Sensitivity analysis.

SYNTHESIS AND DESIGN

A. Copy, Modify, Extend Existing and Successful System

1. Most common, and very useful approach--good initial design

2. Examples: repeated L. C., U.D.C., Dewey, and Uniterm installations

B. Develop from Building Blocks

1. Find all the equipment, tricks, techniques, and gimmicks--pull pieces together in a coordinated manner.

2. Many computer systems are equivalent to manual systems--with a few blocks changed.

3. This approach used by many (architects, mechanical and electrical engineers)--its success depends on the designer's awareness of the availability and capability of hundreds of different kinds of building blocks.

4. Can do some sub-optimization (e.g., minor equipment) and by-pass checking all possible alternatives.

C. Develop a Brand New or "Invented"System

1. System with no obvious ancestral background--very seldom the case.

D. No Conscious Design

1. Systems developed by accident or edict, with no conscious design effort.

EVALUATION

A. Performance Evaluation

1. Single Criterion (pick a single operating parameter--one of many possible ones)

a) Simplest and most common--pose a sample query and look at the volume of responses (a subjective test of adequacy). This does not provide a controlled and critical test--volume per se is not a sufficient measure.

b) Better to compare systems against each other, or against a standard. More useful to have absolute rather than relative measures.

c) Cleverdon's work--UDC, faceted class., conventional subject heading, uniterm, WRU Semantic factors. His test procedure: 1) formulate question from doc. in file; 2) pose this query to the file; 3) measure recall and relevancy.

d) Swanson's and Borko's work to compare manual vs. machine indexing. His test procedure: 1) formulate question from doc. in file; 2) assign a relevance figure to each question--document combination before the test (i.e., a prior determination of all relevant responses prior to the test); 3) pose questions to the file, and score with an expression that penalizes the system for false responses. For example:

$$\text{Score} = \frac{\sum \text{relevance weights of retrieved articles}}{\sum (\text{relevance weights for a given query of all articles in the file})} - \text{Penalty Factor} \left(\frac{\text{number of irrelevant responses}}{\text{number of total responses}} \right)$$

e) Some ~~criterion~~ criticism of synthetic questions.

f) Bornstein suggests two measures: "amount of relevant and non-relevant material recalled: 2) amount of time spent by the user in examining (receiving) the responses.

g) Maizell suggests time as a measure: i.e., what fration of the available ~~tables~~ titles are delivered to the user in a reasonable amount of time?

2. Composite Figure of Merit

a) Quantitatively describe and weight requirements and performance, compute measures of agreement, and obtain a single non-dimensional weighted value.

b) Describe all performance parameters in terms of a common denominator (e.g., time or cost penalties). Search speed and false drop rate both mean something in terms of user's time.

B. Cost Evaluation

1. Can't compute value, but can minimize expenses.

2. Need good cost accounting procedures (standard times and costs). This data is generally unavailable.

3. Can use simulation techniques and detailed econ. analysis to estimate alternate costs of systems during design (paper) stage, over wide variations in operating parameters.

ADDITIONAL COMMENTS

1. Prob. Dfn.--Not done well to date. Techniques are available, just need to put right people on the job (people with the problems are not the people with the tools).
2. Determ. of User Requirements--Tough nut to crack. Needs more work.
3. Synthesis--Arts and Crafts. Need a good knowledge of the bits and pieces.
4. Evaluation--There should be continuing evaluation of any operating library system. The monitoring process should be built into the basic structure of the system. How long since most libraries had a check-up?

Meeting No. 11

COMPUTERS

EQUIPMENT

1. Progression of calculating machines from Babbage (1800's) to desk calculators, to first relay calculators (1950) to stored-program computers. Over 250 different models have been built to date, and over 12,000 are in use today (more computers than special libraries).

2. Computer can be considered the same as a desk calculator, except for two fundamental differences: (1) the instructions can be stored in the machine; (2) the machine can change its program or sequence of operations depending on the results of intermediate calculations.

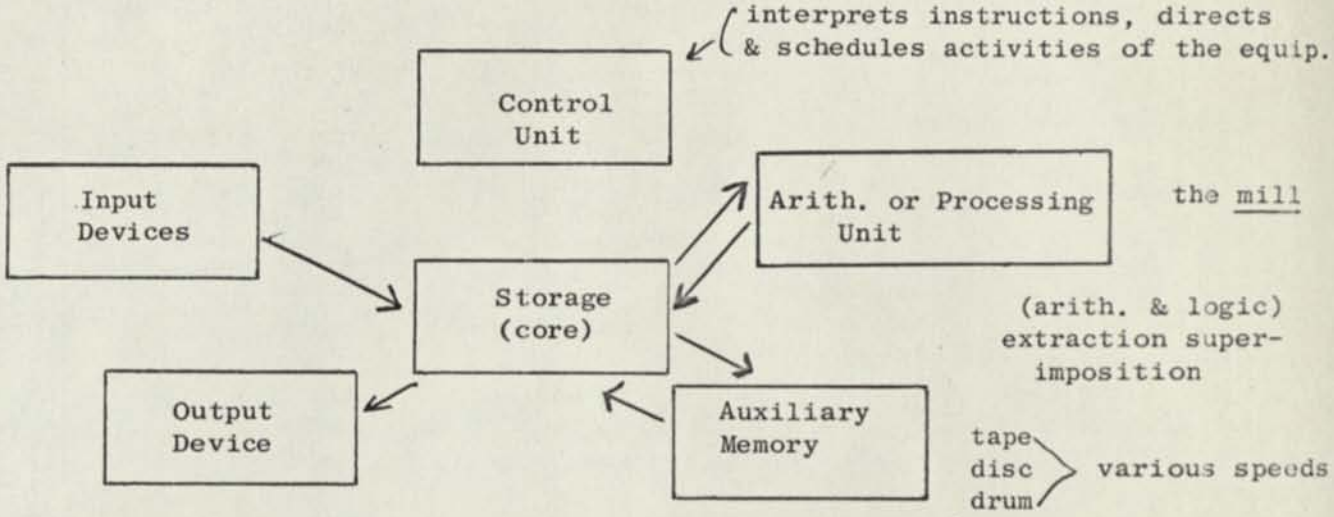
   Example: Functional analysis of printing desk calculator: (1) keyboard one number in the arithmetic register (LOAD, STORE commands); (2) load second number and push MPY button (LOAD, MPY); (3) print result (sequenced automatically by a machine command. The machine also clears the keyboard to allow the next input. The same type of sequence is followed by a computer MPY.

3. Computers are complex-- but you don't have to know how to build one in order to use one (e.g. use of auto without building it). But brief explanation can help provide a better appreciation, and insight of types of problems that might be appropriate for a computer.

4. Can consider the computer to be an extension of tab equipment. It can do the same things (e.g. sorting, merging, selecting, and arithmetic operations)--but at greater speeds, and with greater volumes of data. (Grown-up tab equipment)

5. Sequence of machine operations for each step is still established initially by a human. Thus, these high-speed tab systems are controlled by written lists of sequential machine steps ("programs") rather than by plugboards.

6. Size, shape, and characteristics vary widely--but most computers have some common features (see Fig.)



7. "Random Access" memory does not refer to the way in which information is organized in the memory, nor does it imply immediate access. It means that any memory location, chosen at random, required about the same amount of access time as any other location. (Mag tape is an example of a convenient compromise of cost, capacity, and speed, with up to 50 million digits/reel and 100,000 digits/sec.)

8. Means must be provided for the machine and user to communicate with each other--at such a rate that the machine doesn't have to wait. (p. cards, mag tape, p. tape, MICR, OCR, turn-around documents)

9. Spectrum of problems (scientific vs. data processing problems) e.g. number sieve vs. maintenance of Time-Life subscription list.

10. Some of the equipment (e.g. mag tape) is recognizable, most of it (e.g. arithmetic unit) is not. Operator or Maintenance Consoles are used on every computer to indicate status, and permit some degree of manual control and intermention as required.

PROGRAMMING

1. Computer must be given detailed instructions for every task that it is to perform--just as tab equipment must be given a specified sequence of machine steps by the plugboard wiring--fast mechanical moron. Any credit for cleverness and ingenuity of the computer's operation belongs to the person who prepared the list of instructions (e.g. chess playing or arithmetic programs)
2. Writing the list of instructions is called programming. The programmer works from a specific list of machine operations, and composes a sequence of steps that supposedly foresee all possible circumstances. (e.g. Limit check built in to check over-payment of dividends because of machine errors)
3. Program is a list of separate instructions for elemental tasks. Each instruction is usually represented b a set of numbers or characters arranged in a certain manner, with each number having a pre-determined meaning that is recognized and interpreted by the control circuits.

   Example: (IBM 7070 interrogate command) +5100219999

   "Test tape/disc storage channel #2 to see if it is busy.
   If it is busy, get the next instruction from memory location 9999.
   If it is not busy, execute the next instruction."

This kind of command allows the program to do other chorex while waiting for the busy signal to clear up.

Example: (IBM 7070 tape write command)

+8100551200

"Write, with zero suppression, on magnetic tape transport No. 5 on channel No. 1, the records whose locations are defined by the contents of memory location 1200."

These are "machine language" instructions. The basic language.

4. Problem-oriented languages (programming aids) to make programming easier. (e.g. FORTRAN, ALGOL, COBOL). Statements are now written:

"SORT E ON G

5. Machine languages usually restricted to one model of equipment. Problem-oriented languages may be used by several models (eg RCA, IBM)
6. Program writing is done in a way analogous to report writing: crude outline (crude flowchart), detailed outline (detailed flowchart), completely integrated and coherent text (program steps). Sub-routines (like standard paragraphs) can be borrowed from other sources and patched into the program as required.
7. Extensive programs require a lot of initial effort (e.g. 100 man-years for first SAGE programs). Maintenance programming is also required. It takes time to produce a finished program (iterative debugging)

How to cover this for.

↓

8. Programs can be documented and used by others (program libraries)

APPLICATIONS (next meeting)

1. Analysis of Text Structure
2. Preparation of Retrieval Tools
3. File SEarching
4. SDI
5. Copy Editing
6. Drafting and Related functions

MEETING NO. 12

COMPUTER APPLICATIONS

GENERAL

1. A given problem (e.g., file searching or KWIC) can generally be handled by any of several different models of computers--even though most of the literature shows some predominant types.
2. Use of a computer will not necessarily exhance or improve the efficiency of retrieval--that will depend upon the indexing scheme used. The use of equipment can not provide better retrieval or guarantee good performance, but it might be able to provide more answers in a shorter time or at less cost. The retrieval performance depends upon the index, not on the method of implementation.

ANALYSIS OF TEXT STRUCTURE

1. Concordances. A writing task that has gone over almost completely to preparation by com puters. Definition and description. Uses. It shows how and where each work was used. Take away the accompanying text words, and you have a detailed work index. Take away the locations, and you still have a detailed list of all the words used in the text. Could serve as first draft index or thesaurus. Other possible uses: psycho studies at NIH; interview responses.

   First concordance (Bible) published in 1550, a subsequent effort in 1894 took 30 years of effort.

   First computer concordance (Bible) in 1955 took about 150 hours of computer time (same amount of processing, but shorter elapsed time), with 800,000 words put on 4 reels of magnetic tape. Ignoring 132 junk words, the printed version took 2000 pages.

   Some other words at this time utilized tab cards to make a concordance. (1½ million tab cards.)

   Advantages of computer concordance writing: (1) all mechanical or clerical operations done more effectively, cheaper, with less error, and in shorter elapsed time; (2) output from computer printer can

be used directly for reproduction printer plate--thus eliminating typesetting. The main problem is providing input in machineable form.

Now concordances are being prepared by many people for many different collections of material:

Cornell University Press: Poems of Matthew Arnold (40 hours of 704 time) Anglo-Saxon Poetic Records, Poems of W. B. Yeats, etc.

Dead Sea Scrolls

Penn. legal statutes

Folk songs

Medieval German poetry

2. Dictionaries and Thesauri

Any program that processes text (e.g., KWIC) is in a position to easily obtain data on word usage that can be used to help the development of dictionaries or thesauri . Can monitor dictionaries in the way done for machine translation work. This removes much of the clerical burden from this task.

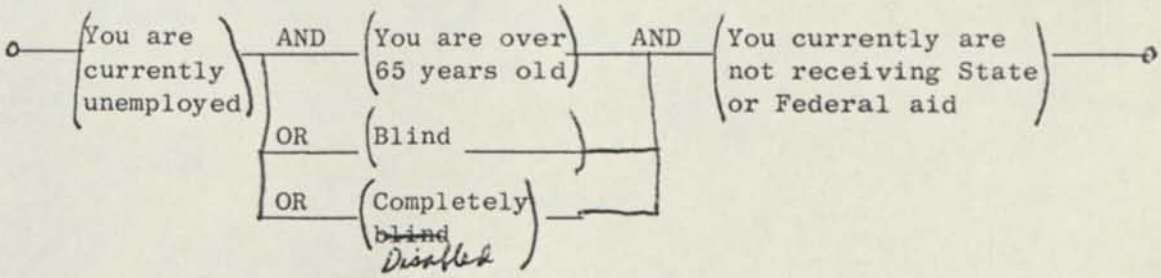
Word frequency counts used to establish authorship of papers (Federalist papers, political ghost writers).

Special dictionaries (e.g., reverse spelling dictionary of 80,000 items at Indiana University).

3. Text Schematics

Some interesting work has been done in analyzing legal statutes, business contracts, and other complex text material by breaking them down into schematic logic diagrams to show their basic meaning.

Example: "You are entitled to County unemployment benefits if:



This is a simple example, but some statutes have sentences that run for a full page. There has been one reported instance of a sentence that was 692 words long.

Consider the possibility of using this approach to generate a "check print" for the person proposing the legislation--and then printed as a guide to accompany the text of the statutes or legal document.

In 1960, 14 million words of Federal statutes were published.

PREPARATION OF RETRIEVAL TOOLS

1. Indexing, Cataloging, and Classification

KWIC has been major application (journal indexes, index journals, indexes to special collections, lists of periodical holdings, etc.) Cost - 1/4 to 1 dollar per bibliographic item to prepare the master repro copy. Examples of formal publication:

KWIC Index to USGRR
Chem Titles
BASIC
Chem Biol. Activities
Index Chemics (rotated or permuted formulas)

KWIC used to permute other things besides citations (e.g., chemical formulas, interview responses). H_2O OH_2

Citation index is another type of machine generated index--still lots of manual effort involved (editing, verification, title checking)

Computer proposed manual research tools:

Uniterm terminal digit cards
Catalog cards
Printed book indexes
Peek-a-boo cards
Tabledex
Scan-Column index

Trend in library use of computers has been to develop a total system that provides many series (e.g., KWIC, catalog cards, accession lists, file searching) at the same time--thus reducing unit costs. Example: Lockheed's 2500 titles/month system provides KWIC and 60-75,000 catolog cards/month, with file ~~recording~~ searching and SDI facilities included.

In all of the above examples, the conputer did only the clerical tasks, and did not do any subject analysis of the material. However, this is still commonly referred to as "machine indexing." The actual attempts to do machine indexing have taken 4 routes:

..check each incoming word against a previously compiled table or thesaurus to achieve word indexing or concept indexing (somewhat inflexible with changes in subject field, notations, etc.). Example: study of a portion of Index Medicus showed that given a subject heading list, and a thesaurus, the computer could make 86% of the assignment the same as human indexers.

..use word frequency statistics (assume that the more often a word appears in a document, the more likely that the word is representative of the subject content of that document. That is, similar word patterns for 2 documents suggest similar contents.

.. Syntactical or grammatical approaches (topic sentences, tip-off words or phrases such as, "In conclusion," prepositional phrases).

..can look at frequency of words occurring close together, comparing these joint probabilities with the average probability of coincidence.

2. Abstracting

Actually extracts rather abstracts. Sentences selected by:

Word frequency and proximity (weighted sentences) topic sentences.

No useful technique in sight for any automatic abstract that actually consists of a re-write or summary condensation.

Cost. $50/4400 word document, which is considerablly higher than present manual cost.

FILE SEARCHING

Computer subject searching of library files started in 1954, using a coordinate index to 25,000 items. Now there are scores of systems in operation (5000 to 100,000 file items). Few of these systems conduct more than a dozen searches/week.

Description of Typical System: file conversion, input and file maintenance, searching, form of output. Response times and delay for record shift time, simultaneous searches with complex logic, no browsing.

G.E. system (90,000 reports): weekly batched run of costs of 8 to 25 cents/biblio item delivered (20 to 120 dollars per customer). Conversion cost and initial programming $20,000.

File search systems also established for searching chemical compounds and their structures.

Coordinate indexes are the most frequent form, but there have been some text searching systems (Describe) Penn. law statutes

ASTIA System: 250,000 reports with 30,000 additions/year. (The index is stored on 4 reels of tape.)
MEDLARS -- in operation next year. Prepare Index Mediums and 50 special biblios/year, plus 100 searches/day.

Trend: Computer tapes shipped to local libraries (e.g., MEDLARS, ASTIA Chemical Patents)

Computers most profitable in situations where:

there is a need for an exhaustive search of a large file;

there is a need for a rapidly response to a query, and the computer is available on demand;

There are complex questions or complex file material (e.g., chemical compounds);

There is a large volume of records
There are no other available bibliographical tools

Problems:

Search at 10,000 items/min., but only once/week in some cases;
Limited printout (number, citation, abstract);
Input cost (perhaps 50 cents/item)
Difficult to provide alternate search strategies or follow associative trails.

SDI

Functional description of operating systems for technical literature (profile matching, feedback, and correction or reinforcement of profile). How do you index or catalog a users' interest profile? All SDI systems do it with key words.

First machine SDI system operated in 1959. Present system now has 1100 clients matched against a stream of 500 documents/month at IBM Lab in Yorktown. New wrinkles: exclude specific journals and other sources at individual user's request.

Random notices get 20-25% response. Test of M.S. degrees vs. high school girls to index showed more responses to indexing of high school girls. Now using high school girls, indexing 15-20 words/document.

Cost: For Yorktown installation (1100 users, 500 documents/month) the cost is about $25/user/year. Acqusition and reproduction of documents is the major cost. Other places have requested $50/user/year.

Trans Canada Airlines using SDI to keep instruction and training manuals up to date (8 mill. pages/year). Match words of manual paragraph headings with words of new orders and directions, and send notices to interested people.

COPY EDITING

1. Copy Generation: List of pronouncable 4-letter words. Programs to generate poems with given ryhme and meter ("you'll wonder where the yellow went...")

2. Copy Editing

   (a) Composition (justification and hyphenation for L.A. Times and others). Paper tape from typewriter is used as input to computer, which then does the processing and punches a new paper tape to drive the typesetting equipment.

   (b) Spelling (misspelled words corrected), spaces inserted when words had been squeezed together). In one instance 4300 words that had been squeezed together were expanded back into text with a 99.7% correction role.

COSTS

1. Rental: $50-500 per hour.
   Purchase: $50,000 to $7 million. (7090 starts $\approx$ $2\frac{1}{2}$ million)
2. Cost of installing and running a computer system (site preparation, training, initial programming, file conversion, parallel testing, etc.) is 2 to 5 times the purchase cost of the computer.

MEETING NO. 13

PAPER TAPE AND MAGNETIC MEDIA EQUIPMENT

PAPER TAPE

1. General. Paper tape often used for temporary storage, point-of-action recording, and input medium for data processing systems.

Popularity due to relatively low cost and simplicity of punching and reading equipment. Tape most normally prepared with an electric typewriter. Can then be used to drive the typewriter.

Methods and problems of error correction. (For errors detected at time of recording, and for errors detected at a later time.)

Tab card punching is faster than paper tape preparation for long jobs or those that require considerable accuracy.

Equipment demonstration. (Flex or Synchrotape or RoboTyper or Auto typist)

2. Application

(a) Duplicate Copies ("personal" letters). Typist types heading and salutation, and enters information where necessary.

Memos

Library Routing Lists

Catolog Cards

Invoice writing (tapes for standard parts and customers)

ASTIA TAB production

(b) Machine Input (for speed and accuracy)

Teletype input (saves extra keyboarding and errors, faster)

ASTIA file input

Cash register transaction recording

Typing of insurance policies provides tape for machine input

(c) File Searching. (WRU only example)

Generally unsiutable: tape wears out, slow mechanical speeds, restricted to serial scanning, awkward to make changes.

3. Cost. $2500-$3200 for Flexowriter.

MAGNETIC TAPE

1. Equipment. Similar in operation to roll magnetic tape, but the magnetic media may take different forms (e.g., Magnacord, CRAM, Videofile).

Advantage: High storage capacity, ease of modification, rapid transfer rate, permanent re-usable storage.

Disadvantages: Data often recorded in serial sequence--with usual problems.

2. Cost. $20,000 to $2 million. None in operational use yet.

Meeting No. 14

MICROFORMS AND MECHANIZED IMAGE SYSTEMS

BACKGROUND

Photographic and microphotographic techniques have been around for over 100 years. Used primarily for security storage (business documents, and then newspapers and other document collections) with low resolution and generally poor quality film and equipment.

Film is still used primarily for security storage--relatively little "publication" in microform. Banks film most of their checks, mail order firms film the orders and return the original papers, other firms keep the film and original records.

Lots of relatively inexpensiae equipment available to do rapid filming of business records, checks, etc. at high speeds (500 checks/min, 1[illegible] pages/min.) and low quality. Equipment and filming for library type film quality costs more. (Flow vs. planetary cameras)

Film sizes: 8 mm and 16 mm common for business records
35 mm most common for libraries and engineering drawings 70 mm and 105 mm sheet film also used.

Reduction ratios: Normally about 20:1 for most applications, but can go 200:1 (diminishing returns)

Resolution: Measure of fidelity. Depends on all the links in the chain. Would like viewer resolution of 7 lines/mm or about 175-200 lines/inch.
(Human eye can resolve 16 lines/mm)

Film Types: silver halide--conventional, with chemical processing

diazo--certain chemicals lose color or are destroyed by light, and can react with ammonia to form dyes. Thus, one can make contact print with light source, and develop with ammonia fumes. Diazo may be impregnated in base, rather than surface coat, so that it is less susceptible to scratching. Cheaper to use than silver film--No need for darkroom.

Kalfax--photosensitive compounds decompose with light, heating converts the decomposed materials to gas bubbles which serve as light scattering centers. Thus contact print is made by exposing film, and then heating (with an iron, boiling water, etc.)--No chemicals needed--and no need for darkroom developing. Add-an-image feature.

Photochromics--Organic dyes that change color with application of a given wavelength of light. Thus can make contact prints with one color light, and erase the image with another color light. Heat and light sensitive. Now used primarily for displays.

FILM VIEWING

Hand viewers -- $50 or less. Generally unsatisfactory. Dozens of models.

Viewer --$300 to $3,000. Getting better all the time

Viewer-printers $700 to $2,000. Getting more use. All film sizes, and paper sizes from 8½ x 11 to 18 x 24

Magazine Viewers -- Recordak--Motorized drive with code marks.

HARD COPY OUTPUTS

Viewer-printers

Copyflo -- 20 ft/min. on 24" paper width (ASTIA, Univ. Microfilm)

Xerox prints for reports, out-of-print books, theses, and dissertations

BASIC MICROFORMS (use examples)

Microfilm roll storage

Conventional (S.F. Chronicle)

With added coding for machine handling (Recordak, FMA, etc.)

Microfilm Unit Record Storage

Aperture cards (any kind and size of stock, any kind of film)

Jackets (various sizes)

Sheet film ("Microfiche")--Actifilm, NASA, etc.

Micro-opaques (Microcard, Microtape)

MECHANIZED IMAGE SYSTEMS

Coding accompanies file item (reel or unit record) vs. addressed storage (reel or unit record)

Coded File Items:

Rapid Selector (1949) developed to FMA Filesearch (1000 ft. rolls with 32,000 frames (roll)--search a roll in 5 minutes. Copy answers onto another film, or view it, or provide display.

Minicard--Unit record (16 x 32mm) with 60:1 reduction. Duplicate copies filled in several reactions of the file. Search at 1200 chips/min.

Filmorex and new Kodak System (Miracode)--Poor man's minicard

Addressed Images:

Lodestar w/ Counting Accessory--Key in frame number for machine search.

CRIS 400 ft. scroll of 17 inches width, stores 500,000 pages. Keyboard designates address for display or aperture card copy. 20 seconds access time.

MEDIA Combined manual-machine system.

WALNUT Film strips in bins (200 cells of 50 strips each, with 99 images on each strap). One unit equivalent to ≈ 3,000 books. Average access time = 5 sec. Output ≈ 500 aperture cards/hr.

APPLICATIONS

Very little application of the large mechanized system, and little likely for conventional library work.

Publication in microform: Still a small activity. 45 publishers with sales $5 mill/year for: (1) copies of material published in other forms (e.g. NY Times Harvard Law Review); (2) material generally unavailable in any other way (e.g. Card Catalog of the U.C. Library, Transcript of Adolph Eichmann Trial, (3) National Bibliographies, historical records); (3) prime form of publication (J. Wildlife Diseases, Tables of Prime Numbers)

Approx. 90% of current U.S. newspapers are available in microform edition, most dissertations available, approximately 30,000 serial titles, manufacturers catalogs. The AEC has distributed over 20 million Microcard copies of its reports.

Cost: Approximately ¼ to ½ cent per original page.

Specific applications: Security storage

Publication

Business and legal records

Medical records

Library--lists of holdings and other catalogs (e.g. Union catalog of National Library of Canada)

Engineering drawings (2 billion dollars/yr. for DOD drawings and reproductions). Files of 50 mill. drawings, with an input of 6 mill. additions or revisions/yr. 28,000 different drawings to describe B-47 aircraft.

F٦

MASTER

Class Problem No. 1 A SAMPLE OF THE COVERAGE OF SECONDARY PUBLICATIONS IN A SCIENTIFIC FIELD

Librarianship 245 (Introduction to Documentation)

Charles P. Bourne, Instructor

Drs. Yariv and Gordon are authorities in the field of optical and microwave applications of solids. They recently published a critical and tutorial review of the field of development of Lasers, a new electronics component. ("The Laser," Proc. IRE, Vol. 51, No. 1, pp. 4-29, January 1963). In this comprehensive review, they cited the original publications that reported many of the valuable contributions in this field. Yariv and Gordon have exercised some value judgement in the selection of their references, so that the references given in their paper can be considered to be important papers in their field.

As a matter of curiosity, and also as a measure of the coverage of the secondary publications, it would be interesting to find out whether these "important" references were ever noted in any of the indexing and abstracting services. The following citations are a few references from this paper. See if you can find these citations in the secondary publications. Note the name, date, and page number, and section heading of the secondary publication in which the article was noted. Make a note of the sources that were examined but did not provide a listing. (As an example, a few notes have been made on the coverage of the first citation.) If some of the original publications seem neglected, while others received coverage by several sources, suggest a few reasons why this might be so. The information from these notes will then be assembled into final form by the Reporter assigned to this particular problem. The report submitted by the Reporter will then be reproduced and distributed to the class members.

Be sure to give appropriate citations, and write your notes in a way that makes them useful for the Reporter.

PAA: Very good idea. Discuss why the secondary publications might not have these citations.

ARTICLE IS CITED IN THESE SECONDARY PUBLICATIONS

Article	Name of Publ.	Date	Page	Section Heading
1. A.L. Schawlow & C.H. Townes, "Infrared & Optical Masers," Phys. Rev. Vol. 112, pp. 1940-1949(Dec. 1958)	Phys. Abstracts	6/59	557	magnetic resonances
	Chem. Abstracts	5/10/59	7765	electronic phenomena
2. T.H. Maiman, "Stimulated Optical Radiation in Ruby Masers," Nature, Vol. 187 pp. 493-494 (Aug. 1960).				
3. A. Javan, W.B. Bennett, Jr. and D.R. Herriott, "Population Inversion and Continuous Optical Maser Oscillation in a Gas Discharge Containing a He-Ne Mixture," Phys. Rev. Lett. Vol. 6, pp. 106-110(Feb. 1961)				
4. L.F. Johnson, et al "Continuous Operation of the $CaWo_4:Nd^{3+}$ Optical Maser," Proc. IRE (Correspondence) Vol. 50, p. 213 (Feb. 1962).				
5. D.F. Nelson and W.S. Boyle "A Continuously Operating Ruby Optical Maser," Appl. Optics Vol. 1, pp. 181-183 (March 1962).				

THESE SECONDARY PUBLICATIONS WERE CHECKED BUT DID NOT HAVE THE CITATION

Name of Publ.	Date
Electrical Eng. Abstracts	1959,1960

Librarianship 245: (Intro. to Documentation)

CLASS PROBLEM NO. 1 (A SAMPLE OF THE COVERAGE OF SECONDARY PUBLICATIONS IN A SCIENTIFIC FIELD.)

R. Siebert, Reporter

Objective:

To analyze the coverage of secondary publications (abstracting and indexing media) in regard to a selected sample (tracers) of articles appearing in scientific publications.

Given:

Five articles that reported most of the new and original contributions in the field of Lasers, a new electronic component.

Problem:

Determine to what extent these important and original contributions were reported in the indexing and abstracting publications. Comment on why they might not all receive the same degree of coverage.

Abbreviations used in this Report:

ASTI	Applied Science and Technology Index
CHEMAB	Chemical Abstracts
CHEMTI	Chemical Titles
EEAB	Electrical Engineering Abstracts
ENGIND	Engineering Index
NUSAB	Nuclear Science Abstracts
PHYSAB	Physics Abstracts
SSAB	Solid State Abstracts

Analysis:

1. Coverage by Secondary Sources

Secondary Source	Number of these 5 articles that were picked up by this secondary source	No. of the 5 different article sources that are "covered" by this secondary sourc	
Physics Abstracts	4	4	(Applied
Chemical Abstracts	2	3	Optics missing
Chemical Titles	1	3	
Electrical Engineering Abstracts	1	4	
Engineering Index	1	3	
Solid State Abstracts	1	2	or more
Nuclear Science Abstracts	none	2	
Applied Science & Technology Index	none	1	
British Science & Technology Index	none		
Math Reviews	none		
Reader's Guide to Periodical Lit.	none		
Bibliographie der Fremds. Z. Lit.	none		

The factors operating to produce this pattern are:

1. the particular coverage of each secondary publication. The scores change somewhat when we note that:
   a. ASTI only indexes the IRE Proc. score: 0
   b. CHEMAB indexes Nature, Applied Optics, and Phys. Rev. 66-2/3
   c. CHEMTI indexes Nature, Phys. Rev., and Phys. Rev. Letts. 33-1/3
   d. EEAB indexes Nature, Phys. Rev., and Phys. Rev. Letts. as well as IRE Proc. 25
   e. ENGIND indexes Nature, Phys. Rev. and IRE Proc. 33-1/3
   f. NUSAB indexes Nature and Phys. Rev. 0
   g. PHYSAB does not index Applied Optics 100
   h. SSAB it could not be determined what journals this publication covers except for Phys. Rev. and Applied Optics.

2. the time lag before articles are noted in the secondary publications. Perhaps the coverage score may be raised when the semi-annual and annual editions for 1962 and 1963 are published.

3. careless checking by the students. For instance, two articles in ASTI and one in ENGIND picked up by students were not in the primary journal cited but in another journal under the same title and author. Another title, picked up in CHEMTI had a similar title, but was not the same. Obviously, searching stopped when these "bogies" were found; if continued, the proper articles might have been found, thus raising the score of these publications.

Analysis:

2. Nature of the coverage of each article

Primary Journal Source of Articles	Number of Secondary Journals that Index this Primary Journal
Nature (Ref. 2)	6
Phys. Rev. (Ref. 1)	5
Phys. Rev. Letts. (Ref. 3)	4
IRE Proc. (Ref. 4)	3
Applied Optics (Ref. 5)	2

It might be concluded that Applied Optics is deemed to be too specialized for the editors of CHEMTI, ENGIND, NUSAB, EEAB, ASTI and PHYSAB. The correspondence of the IRE Proc. is probably ignored as not valuable enough to bother abstracting.

Analysis:

3. Time Lag between Primary Publication & Secondary Coverage

	Elapsed time before being picked up (in months)				
Secondary Journal	Article No. 1	2	3	4	5
PHYSAB	6	4	4	4	
CHEMAB	5		10		
ENGIND	12				
CHEMTI			1		
EEAB				7	
SSAB					2

A substantial lag exists between the time when an article is published and it appears in the indexing and abstracting media. For the sample selected, the average time lag was 6 months. It may be many more months before the articles not picked up are indexed. This presents a tremendous problem for scientists who need to know almost immediately what is being done in their special fields. If they do not catch the article in the primary journal in which it is published, they cannot depend on the bibliographic journals to give them current information.

Analysis:

4. Use of Subject Headings in the Secondary Publications

No two secondary journals placed the articles under the same or similar headings. This presents the user with a difficult problem when it comes to literature searches.

ARTICLE IS CITED IN THESE SECONDARY PUBLICATIONS.

Article	Name of Publ.	Date	Section Heading Used
1. A.L. Schawlow & C.H. Townes, "Infrared & Optical Masers," Phys. Rev. Vol. 112, pp. 1940-1949(Dec.1958).	PHYSAB	June, 1959	Magnetic Resonances
	CHEMAB	May 10, 1959	Electronic Phenomena
	ENGIND	1959	Masers
2. T.H. Maiman, "Stimulated Optical Radiation in Ruby Masers," Nature, Vol. 187 pp. 493-494 (Aug. 1960).	PHYSAB	Dec., 1960	Optical Properties of Solids
3. A. Javan, S.B. Bennett, Jr. and D.R. Herriott, "Population Inversion and Continuous Optical Maser Oscillation in a Gas Discharge Containing a He-Ne Mixture," Phys. Rev. Lett. Vol. 6, pp. 106-110(Feb.1961).	PHYSAB	June, 1961	Atoms. Isotopes.
	CHEMTI	March 5, 1961	Oscillation
	CHEMAB	Dec. 11, '61	Electronic Phenomena and Spectra
4. L.F. Johnson, et al "Continuous Operation of the $CaWo_4:Nd^{3+}$ Optical Maser," Proc. IRE (Correspondence) Vol. 50, p. 213 (Feb. 1962).	EEAB	Sept.,1962	Masers. Lasers
5. D.F. Nelson and W.S. Boyle "A Continuously Operating Ruby Optical Maser," Appl. Optics Vol. 1, pp. 181-183 (March 1962).	PHYSAB	July, 1962	Electromagnetic waves and oscillators
	SSAB	V.3, no. 4, '62	Paramagnetic devices (Masers and Lasers)

THESE SECONDARY PUBLICATIONS WERE CHECKED TO THE LATEST ISSUE, BUT DID NOT HAVE THE CITATION.

Name of Publication	
EEAB	
ASTI	
EEAB	NUSAB
SSAB	ENGIND
ASTI	
CHEMAB	
CHEMTI	
EEAB	
SSAB	
ASTI	
NUSAB	
ENGIND	PHYSAB
SSAB	
CHEMAB	
NUSAB	
ASTI	
CHEMTI	
EEAB	
CHEMTI	
ASTI	
ENGIND	
CHEMAB	
NUSAB	

Class Problem No. 2 A SAMPLE OF THE ECONOMICS OF MICROFORM STORAGE
Librarianship 245 (Introduction to Documentation)
Charles P. Bourne

The director of a large library was faced with the following question to resolve: Should he bind and store the 1962 issues of Physical Review, or should he purchase and store a microform edition of that periodical and throw away the original paper copies? From a cost viewpoint, which is the more favorable alternative?

To help your analysis, make the following assumptions:

1. Storage space in the library costs $5.00 per linear foot of shelving per year.
2. The cost of binding would be $40.00 for the 1962 issues, and would be uniformly distributed over the number of years that the material is to be stored.
3. The initial cost of the Microform edition can be obtained from A. J. Diaz, Subject Guide to Microforms in Print, 1962-1963 (Microcard Editions, Inc., Wash. D.C., 1962), and would be uniformly distributed over the number of years that the material was to be stored.
4. The cost of the microfilm storage would be negligible. **

State and use any other assumptions that you feel are necessary. Plot your cost data on the graph below, and attach your calculations on a separate sheet.

What other cost factors might be considered in choosing between these alternatives? What other factors might be considered in addition to the relative economics of the alternatives? *

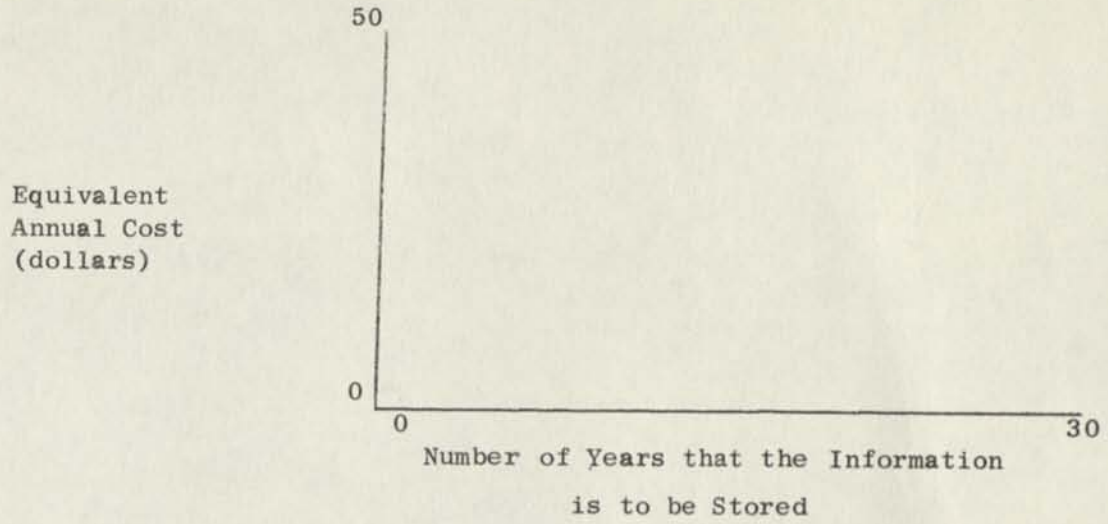
The information from these summary notes will be assembled into final form by the Reporter assigned to this particular problem. The report submitted by the Reporter will then be reproduced and distributed to the class members.

? Is this amortizing the cost of the building?

5 v. in 10 This price is low

* The faculty would rather drop dead than use microfilm. Nobody, but nobody, can tell them they have to! Cost factor not considered significant!

** Don't forget the readers + the hard copy reproducer. Also Microtext Room is in another building from Phys. Rev.



Suggested Reading

E. Power, "Microfilm--The Versatile Academic Tool," in Microtexts as Media for Publication (hertfordshire County Council Tech. Info. Service, Hertfordshire, England, 1960).

E. Power, "Microfilm as a Substitute for Binding," American Documentation Vol. 2, (1951).

Librarianship 245: (Intro. to Documentation)

CLASS PROBLEM NUMBER 2.
F. Siemon, Reporter

Introduction

This report summarizes the conclusions made by sixteen class members in Class Problem Number 2: A Sample of the Economics of Microform Storage.

Problem

The class was to calculate the relative costs of three alternative methods of storing the 1962 issues of the journal Physical Review and to recommend the use of one method which would yield the lowest equivalent annual cost* for various periods from one to thirty years. Data was to be reported in graphic and tabular form showing the relative costs for each of the three methods as a function of the length of time that the material was to be stored. Students were to report other factors (aside from that of cost) which might affect the choice of method.

The three methods considered for storing the journal were: (1) storing the original unbound copies; (2) binding and storing the original copies; (3) discarding the original copies and purchasing a microform edition of the journal.

For this problem the following simplifying assumptions were made: (1) that storage space in the library costs five dollars per lineal foot of shelf space per year of storage; (2) that binding cost is forty dollars which is uniformly distributed over the number of years the journal is to be stored; (3) that the cost of microform storage is assumed to be negligible; (4) that reading equipment for microforms is already available.

Procedure

The per year cost of just storing the original copies of the journal in bound and unbound form was found. In this case, the unbound journal issues required approximately 1-1/2 feet of shelf space resulting in an annual storage cost of approximately $7.50 while the bound journals required approximately 1-3/4 feet of shelf space resulting in an annual storage cost of $8.75. From A.J. Diaz, Subject Guide to Microforms in Print, 1962-63, the purchase cost of the microform edition of Physical Review was found to be $25.45.

The equivalent annual cost for the first alternative (i.e. storing the unbound journals) was simply the per year cost of storing them. This is expressed by:

EAC_1 = Annual storage cost = 7.50 dollars.

*For the purposes of this report, equivalent annual cost is defined as the total cost of storing the material for a given number of years, divided by the number of years the material is stored.

The equivalent annual cost for the second alternative (i.e. storing the bound copies of the journal) was simply the annual storage cost of the bound copies plus the binding cost distributed uniformly over the number of years (N) the material was stored. This is expressed by:

$$EAC_2 = \text{Annual storage cost} + \frac{\$40.00}{\text{Number of years material was to be stored}} = \left(8.75 + \frac{40}{N}\right) \text{ dollars}$$

The equivalent annual cost for the third alternative (i.e. storing the microform copy) was simply the purchase cost of the microform copy distributed uniformly over the number of years stored. This is expressed by:

$$EAC_3 = \frac{\text{Purchase cost of the microform edition}}{\text{Number of years material is stored}} = \frac{25.45}{N} \text{ dollars}$$

Results

The equivalent annual costs of the three alternatives are shown in the following Table and in Figure I. It can be seen that the first alternative, i.e. storage of unbound copies is the cheapest method if the material is to be stored less than about four years. Beyond that time however, the third alternative (i.e. microform storage) becomes the cheapest method of storing the 1962 issues of the journal Physical Review.

Table I. Equivalent annual costs for three alternative methods of storing the 1962 issues of Physical Review for N years.

Number of years (N)	1. Unbound Storage	2. Bound Storage	3. Microform Storage
1	7.50	48.75	25.45
2	7.50	28.75	12.75
3	7.50	22.08	8.48
4	(Constant)	18.75	6.36
5		16.75	5.09
6		15.42	4.24
7		14.46	3.64
8		13.75	3.18
9		13.19	2.83
10		12.75	2.55
15		11.52	1.70
20		10.75	1.28
25		10.35	1.02
30		10.08	.85

Conclusion

The microform approach was found to be the cheapest method of storing the 1962 issues of the journal Physical Review in the event that it was to be stored for more than four years. However, this conclusion was based on a very simple analysis which did not take account of such considerations as the cost of purchasing reading equipment, and the maintenance such equipment might require. This conclusion also ignores such intangibles as the possible user dissatisfaction with materials in microform.

PROB. NO. 2

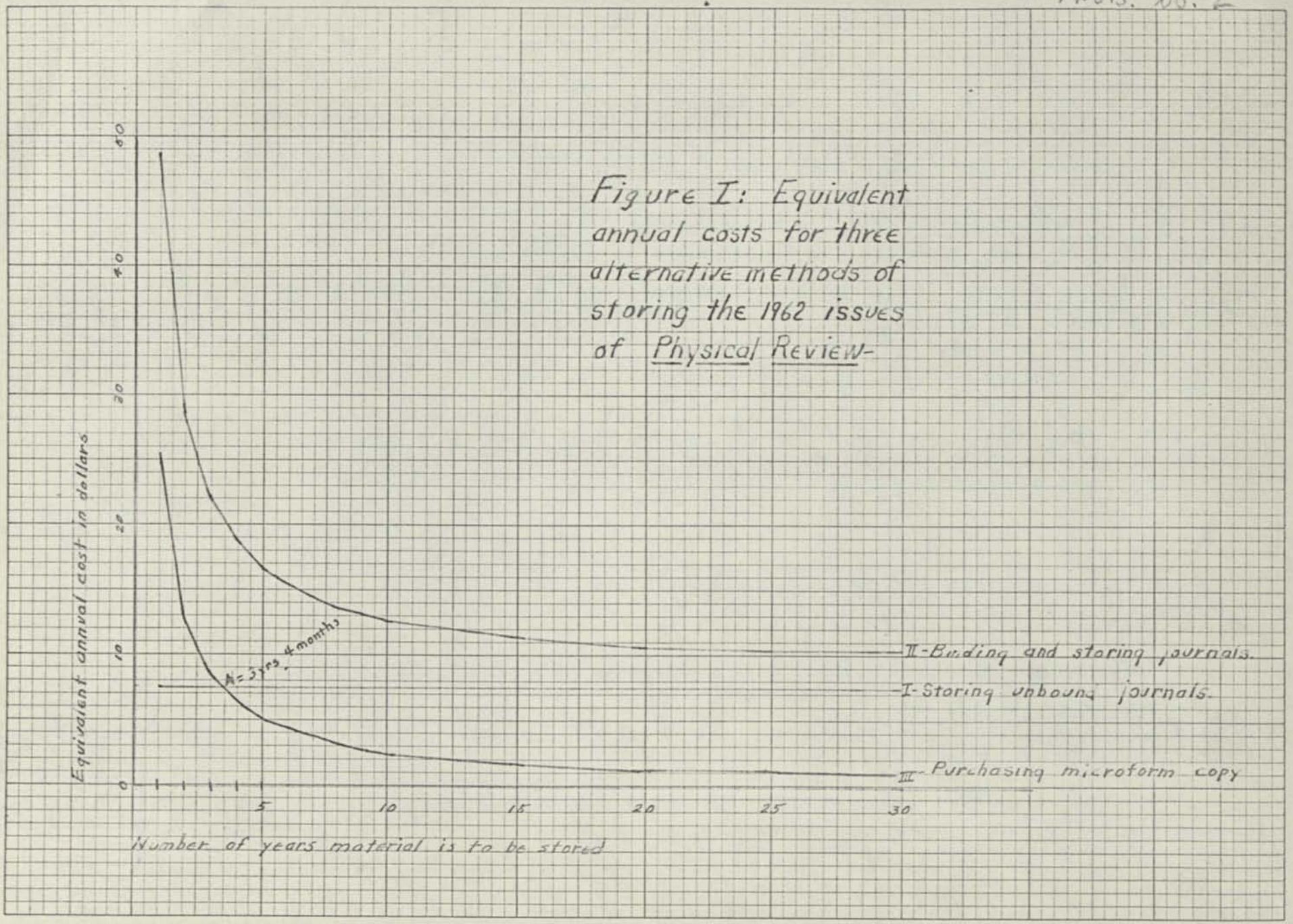


Figure I: Equivalent annual costs for three alternative methods of storing the 1962 issues of Physical Review-

Equiv. annual cost

Plan A

Find & Store

Binding cost = $40.—

annual Storage cost = $5./ft

~~Sing~~ Total Shelf Space Required? S

~~Annual~~ Storage cost = S ft @ 5.00/ft/yr = N^{yrs} 5.00 SN

annual portion

For 10 years storage, total cost = $\frac{40 + \text{Storage/yr.}}{N}$ ~~10~~ N

Plan B

Buy microform from Univ. Microfilm

25.45

Cost/yr. = $\frac{25.45}{N}$

No. yrs	Cost/yr.
1 yr.	= 25.45
2 yr.	= 12.72
3 yr.	= 8.48
4	= 6.36
5	= 5.09
6	= 4.24
7	= 3.63
10	= 2.54
30	= .85

Plan C

No binding – Store in original form

annual cost = constant = ?

Class Problem No. 3 THE ECONOMICS OF FULFILLING REQUESTS FOR OUT-OF-DATE JOURNAL ARTICLES

Librarianship 245 (Intro. to Documentation)

Charles P. Bourne, Instructor

Many studies have shown empirically that the number of requests for articles from almost any scientific journal * are a function of the age of the article, and generally fall off rapidly after the original date of publication. You might say that, on the average, the value of any journal article has a "half-life" in a manner analogous to the half-life of radioactive compound.

The director of one large library hoped to be able to take this behavior into account in designing the procedures for the library to follow in satisfying user requests for journal articles. This director found that for a particular chemical journal (1 annual vol. of 12 issues), the usage followed the pattern described in the table below.

Year After Publication (N)	Number of Requests for Articles (from the entire vol.) During Nth Year after Publication
1	64
2	32
3	16
4	8
5	4
6	2
7	1
8	1
9	1
10	1

This means that there were 64 requests for articles from the entire journal volume where the complete volume was less than one year old, and that there were 2 requests for articles from this volume when it was between 5 and 6 years old.

The director has several means for providing the service, and each alternative has an associated cost. The problem he faces is to find the most economical approach, or combination of approaches, for providing this

* not true in humanities or social sciences

service that falls into the pattern indicated by the previous table. For this exercise, assume that the director has just handed this task to you, so that you must find the answer for him.

Assume also that the library has some special operating rules: (1) one copy of the journal will be subscribed to in any case (thus the subscription cost is common to all the alternatives and need not be considered in the cost comparisons); (2) the original journals need not be kept in the file after all 12 issues have been received;* (3) requests are always fulfilled by providing page-size hard copies of the originals, since the journals (in paper or microform) are not allowed to leave the stacks.

The most obvious ways to provide the service are the following:

1. Store the journals in their original form. Their storage cost is estimated to be $4.00 per year. Hard copy can be prepared from office copying equipment at a cost of ~~$0.40~~ per request.

2. Bind and store the journals in their original form. The binding cost is estimated to be $15.00. The storage cost is estimated to be $3.50 per year. Hard copy can be prepared from office copying equipment at a cost of $0.40 per request.

3. Obtain a microform copy of the journal at a cost of $18.00. Hard copy can be made on a microform viewer-printer (already installed in the library) at a cost of $0.50 per request. The storage cost is negligible.

4. The material can be obtained on inter-library loan from a neighboring library at a total cost of $5.00 per request.

Assuming that any of these routes, or combinations thereof, provide the service in the same amount of time, and with the same quality of hard copy, what plan of operation would you propose? What would be the total cost for providing the service with your proposed plan? What other factors (economic or otherwise) might be considered in making a decision about these alternatives? What are the arguments for binding the journals instead of storing them in their original form?

* What does this mean - no binding?

State and use any other assumptions that you feel are necessary. Attach your calculations on a separate sheet. The information from these summary notes will be assembled into final form by the Reporter assigned to this particular problem. The report submitted by the Reporter will then be reproduced and distributed to the class members.

Suggested Reading: P. A. Richmond, "A Proposal for Dual Publication of Scientific Journals," American Documentation, Vol. 14, No. 1, pp. 54-55 (January 1963).

Librarianship 245: (Intro. to Documentation)

CLASS PROBLEM NUMBER 3.
F. Siemon, Reporter.

Introduction

This report summarizes the conclusions made by sixteen class members in Class Problem Number 3: The Economics of Fulfilling Requests for out-of-date Journal Articles.

Problem

The class was to calculate the relative cost of four alternative methods of fulfilling requests for a popular journal. Recommendations were to be made as to which alternative should be chosen on the basis of which method resulted in the lowest total cost for a ten year period. Students were to report other factors (aside from that of cost) which might affect the conclusions drawn.

The four alternatives initially suggested were:

(1) Store the journal in unbound form at a storage cost of $4.00 per year, providing copies of requested articles by using library copying equipment at a cost of $0.40 per request.
(2) Bind and store the journal. The binding would cost $15.00; the storage cost would now be $3.50 per year, and copies of requested articles would be provided by using copying equipment at a cost of $0.40 per request.
(3) Obtain a microform copy of the journal at a cost of $18.00, from which copies would be made at a cost of $0.50 per request. The library would discard the original copies after the microform copy had been obtained.
(4) Discard the original copies of the journal and fulfill requests from interlibrary loan at a total cost of $5.00 per request.

The type of journal in question was found to have a usage pattern of the form shown below:

Year after publication.	Number of requests for the year.
1	64
2	32
3	16
4	8
5	4
6	2
7	1
8	1
9	1
10	1

For this problem the following simplifying assumptions were made: (1) that service provided by the four methods was equal in all respects save cost; (2) storage cost of the microform was negligible.

Results

The following table shows the cost per year and total cost reported for each of the following methods:

Year	(1) Unbound	(2) Bound	(3) Microform	(4) Interlibrary Loan
1	$29.60	$30.60	$33.80	$320.00
2	16.80	17.80	17.80	160.00
3	10.40	11.40	9.80	80.00
4	7.20	8.20	5.80	40.00
5	5.60	6.60	3.80	20.00
6	4.80	5.80	2.80	10.00
7	4.40	5.40	2.30	5.00
8	4.40	5.40	2.30	5.00
9	4.40	5.40	2.30	5.00
10	4.40	5.40	2.30	5.00
Total Cost:	$92.00	$102.00	$83.00	$650.00

Conclusion:

Of the four alternatives initially suggested, the third method (i.e. discarding the original issues of the journal and purchasing the microform edition) was the best alternative, being less expensive than all other three methods.

However, six students found that a fifth alternative which was a combination of methods 1 and 3 (i.e. using unbound journals for one year and then discarding them and purchasing a microform edition) would result in a total cost of $80.60 for the ten year period. This fifth alternative was the most attractive scheme proposed. Two students found that a sixth alternative which was a combination of storing the unbound journals for two years and then buying the microform edition of the journal, would result in a total cost of $81.40 for the ten year period.

In actual practice, the selection of alternatives would not be as simple as this economic model, and would consider other costs (e.g. storage and maintenance of the microform viewer) and intangibles (e.g. user reaction to microforms).

1. Store orig. smoke detect Cost = $4.00/yr. + 0.40/report

Year	Total Cost
1	4. + 25.60 = 29.60
2	4 + 12.80 = 16.80
3	4 + 6.40 = 10.40
4	4 + 3.20 = 7.20
5	4 + 1.60 = 5.60
6	4 + .80 = 4.80
7	4 + .40 = 4.40
8	4 + .40 = 4.40
9	4 + .40 = 4.40
10	4 + .40 = 4.40
	$92.00

2. Bird + store orig. cost = $15. binding + ~~3.00~~ 3.00/yr. + 0.40/report

Year	Total Cost	
1	15 + 3.00 + 25.60 = ~~44.10~~	43.60
2	3.00 + 12.80 = 16.30	15.80
3	3.00 + 6.40 = 9.90	9.40
4	3.00 + 3.20 = 6.70	6.20
5	3.00 + 1.60 = 5.10	4.60
6	3.00 + .80 = 4.30	5.80
7	3.00 + .40 = 3.40	3.40
8	3.00 + .40 = 3.40	3.40
9	3.00 + .40 = 3.40	3.40
10	3.00 + .40 = 3.40	3.40
	102.00	97.00

3. microform; thousands orig. Cost = $18. + $0.50/report

Year	Total Cost
1	18 + 32 = 50
2	16
3	8
4	4
5	2
6	1
7	.50
8	.50
9	.50
10	.50
	83

4. I-L Cost = 5.00/trans

Yr.	Total Cost
1	320
2	160
3	80
4	40
5	20
6	10
7	5
8	5
9	5
10	5
	650.

5. Straight line method for 1 yr., Uniform Gradient

Yr.		Total Cost
1	4125.60	29.60
2	18+16	34.
3		8
4		4
5		2
6		1
7		.50
8		.50
9		.50
10		.50
		80.60

6. Straight line method for 2 yr. - Uniform Gradient

Yr.		Total Cost
1		29.60
2		16.80
3	18+8	26. -
4		4
5		2
6		1
7		.50
8		.50
9		.50
10		.50
		81.40

Steve orig. & subs kept
Bird to Steve ...
Microform
Ill

	Steve orig. & subs kept	Bird to Steve orig. & subs kept	Hard copy from microform thru orig	Interlibrary loan	Steve orig. & kept for 1 yr., then to microform	Steve orig. & kept 2 yrs., then to microform
1	29.60	43.60	50.	320	29.60	29.60
2	16.80	15.80	16	160	34.	16.80
3	10.40	9.40	8	80	8	26.
4	7.20	6.20	4	40	4	4
5	5.60	4.60	2	20	2	2
6	4.80	3.80	1	10	1	1
7	4.40	3.40	.50	5	.50	.50
8	4.40	3.40	.50	5	.50	.50
9	4.40	3.40	.50	5	.50	.50
10	4.40	3.40	.50	5	.50	.50
	92.00	97.00	83.00	650.	80.60	81.40

Class Problem No. 4 TIME LAGS IN PRIMARY PUBLICATIONS
Librarianship 245 (Intro. to Documentation)
Charles P. Bourne, Instructor

For the rapid and timely communication of research results and new ideas, it is important that an author's work be published relatively soon after the completion of his work. Some delays are unavoidable. It takes a finite amount of time to review, edit, and publish an article--regardless of the type of publication. One of the most common means for the first reporting of original research results is publication in one of the primary journals. However, there has been a considerable amount of criticism of these journals, mostly directed at their long time lags between the time that a manuscript is submitted to the publisher, until the time that the work is actually published. It is the intent of this exercise to determine the nature of these time lags for a number of representative publications.

Each of the journals listed below follows an editorial practice of publishing a note with each article to show when the author's final manuscript was received for publication. A comparison of this figure with the date of the publication will tell the elapsed time from final submission until publication. (This assumes that the journal was published on time.) Each student will be assigned a different journal to study. For the journal assigned, analyze the 50 most recent articles to determine their time lag (to the nearest month) in the manner defined above. Record the requested data in the spaces provided at the end of these notes.

The information from these summary notes will be assembled into final form by the Reporter assigned to this particular problem. The report submitted by the Reporter will then be reproduced and distributed to the class members.

JOURNALS TO BE STUDIED

Amer. J. Physics

Applied Physics Letters

ARS Journal

Astrophysical J.

British J. Applied Physics

British J. Radiology

Bull. Amer. Assoc. of Petroleum Geologists

Canadian J. Physics

J. Acoustical Soc. of America

J. Aerospace Sciences

J. Amer. Ceramic Soc.

J. Amer. Chem. Soc.

J. Biological Chemistry

J. Chemical Physics

J. Economic Entomology

J. Fluid Mechanics

J. Mathematical Physics

J. Mechanical Eng. Science

J. Organic Chemistry

J. Physical Chemistry

Physical Review

Physical Review Letters

Trans. Faraday Soc.

Student's Name _____

Name of Primary Journal Studied: _____

Frequency of Publication: _____

Dates of journal issues that included the 50 chosen articles _____

to _____

No. of articles with this time lag (in months):

1 2 3 4 5 6 7 8 9 10 11 12

13 14 15 16 17 18 19 20 21 22 23 24 >2 yrs.

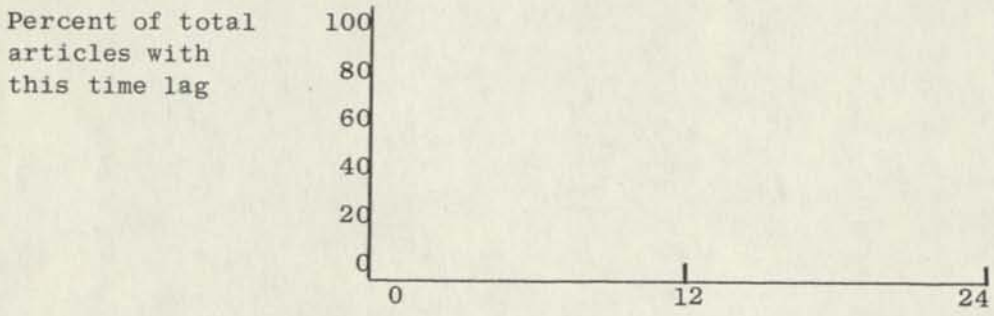
Percent of total (50) articles with this time lag (in months):

1 2 3 4 5 6 7 8 9 10 11 12

13 14 15 16 17 18 19 20 21 22 23 24 >2 yrs.

Average time lag:

Plot your percentage data and average on the chart below:



Approx. amount of time spent on this exercise: _____

Comments on procedures that would simplify this exercise:

3 April 1963

Class Problem No. 4 TIME LAGS IN PRIMARY PUBLICATIONS
Librarianship 245 (Intro. to Documentation)
Class Reporter: Charles Shain

1. Introduction

Prompt communication of research findings and insights is critical in today's scientific world in order to reduce expensive work duplication and advance the solution of problems. Primary publications (called "journals" in this paper for brevity) publish original papers and are an important means to circulate new findings. However, these journals are often criticized for their delays in publication of such papers, beyond the date when they have been fully edited and are ready for publication.

It is the purpose of this exercise to examine the length of this delay, called the "time lag", in 16 representative journals, to determine whether any recognizable pattern exists. The journals in question print the date that each fully edited manuscript was received, and this date will be compared to the formal publishing date of the journal (assuming the journal was published on time). Time intervals less than one month have been disregarded, and are given to the nearest month.

Each student selected one journal from a list furnished for this exercise, and examined 50 articles appearing in the latest available issue(s) of this journal. The "time lag" was calculated for each article, and an "average time lag" (the arithmetic average) was calculated. These individually calculated values and other data were then examined and compared by the class reporter. These values, and the Reporter's interpretation are reported in this note.

2. Findings. The following delays were determined:

Frequency of Publication	Journal	Range of Delay for Individual Articles (mos.)	Average Delay for 50 recent articles(mos.)	Average Delay for this Group (mos.)
Twice each mo.	Physical Review Letters	1-5	1.54	
"	J. Amer. Chem. Soc.	3-9	5.95	4.55
"	Physical Review	5-10	6.16	
Monthly	Appl. Phys. Letters	2-5	2.90	
"	J. Aerospace Sci.	2-12	4.00	
"	J. Organic Chem.	4-26	5.00	
"	Amer. Rocket Soc. J.	1-10	5.00	
"	Canad. J. Physics	3-11	5.02	
"	J. Acoustical Soc.	3-13	5.64	
"	J. Biolog. Chem.	3-12	5.96	6.00
"	J. Phys. Chem.	5-7	6.46	
"	Amer. J. Physics	3-16	6.56	
"	J. Physical Chem.	4-11	7.44	
"	J. Math. Phys.	4-20	8.00	
"	Bull. Am. Assoc.Petrol. Geol.	2-17	8.76	
8 times each yr.	Astrophysical J.	5-11	6.58	6.58
Every 2 months	J. Econ. Entomology	3-11	7.74	7.74
Quarterly	J. Mech. Eng. Sci.	3-13	6.92	6.92

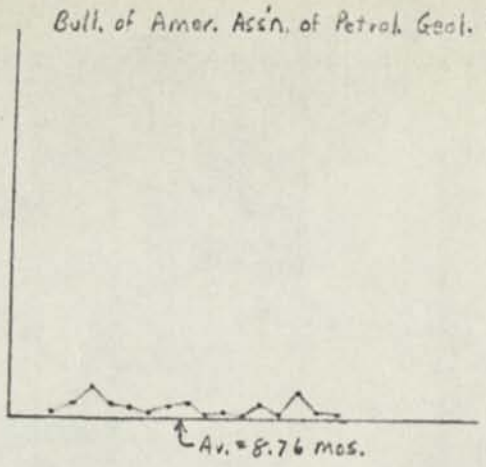
3. Interpretation.

In each publishing schedule the sample sizes are too small to permit generalizations about the delays expected for journals with different frequencies of publication. There are only single instances of journals which exceed the interval of monthly publication, so that it would be difficult to determine an "average" value in these cases. Nevertheless, they all exceed the "average" time lag of the semi-monthly and monthly publications -- 7 to 8 months, compared to 4 to 6 months for the more frequent journals. Yet, the highest "average" time lag in the monthly magazines, almost 9 months, exceeds their range. This precludes an automatic identity of shorter publishing schedule with shorter time lags.

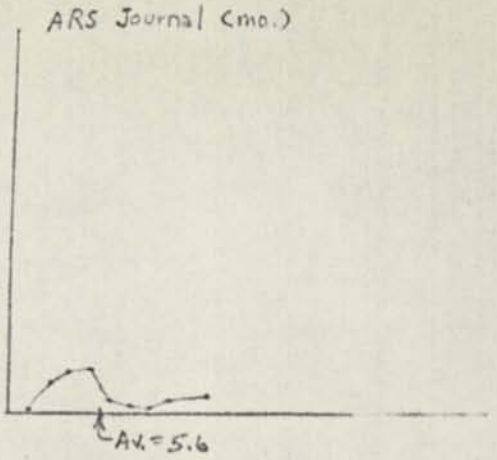
Physical Review Letters and Applied Physics Letters. The one clear pattern occurs in connection with the two "letters" journals vs. all other journals in the semi-monthly and monthly categories. A definition of their distinguishable editorial policy would seem to include the concept of publishing brief communications, or articles (hence "letters"), to speed up the communication of scientific findings.

Disregarding the small size of this sample - in particular the single examples for the "letters" journals, and the other single semi-monthly journal -it appears that the editorial response time of the monthly "letters" journal exceeds that of the other semi-monthly journal by 3 months. If this pattern were to continue in larger, more representative samples, serious consideration of this "letters" editorial policy - in terms of broader application - would certainly be in order. It may be that other factors might come into play, cancelling out some of the advantages of brief, relatively rapid communication. This could result in a new publication policy "mix", to the advantage of the communication of scientific papers.

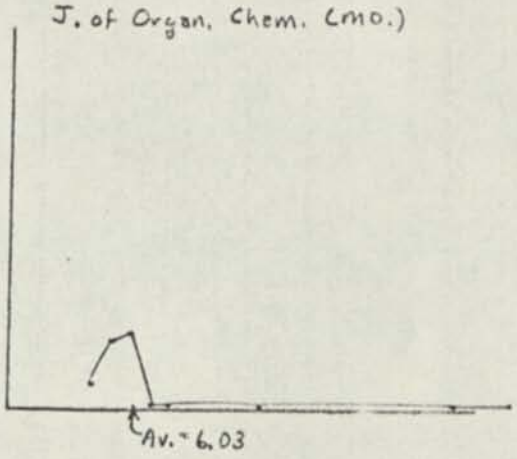
Bull. of Amer. Ass'n. of Petrol. Geol. (mo.)



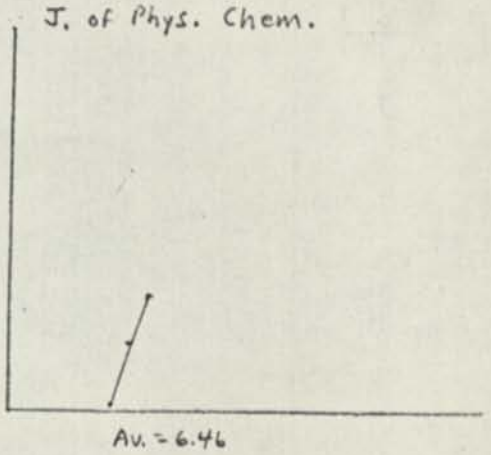
ARS Journal (mo.)



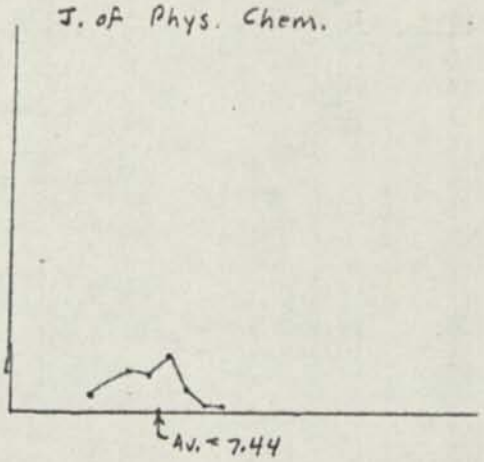
J. of Organ. Chem. (mo.)



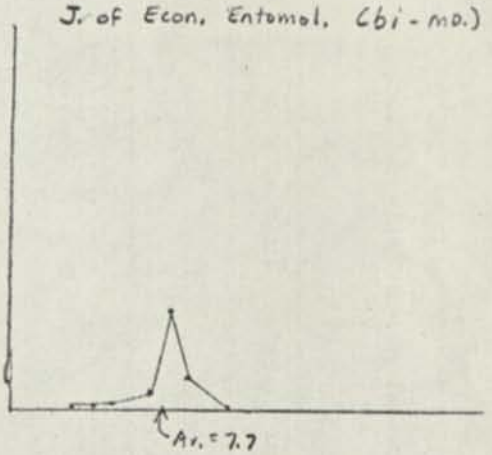
J. of Phys. Chem. (mo.)



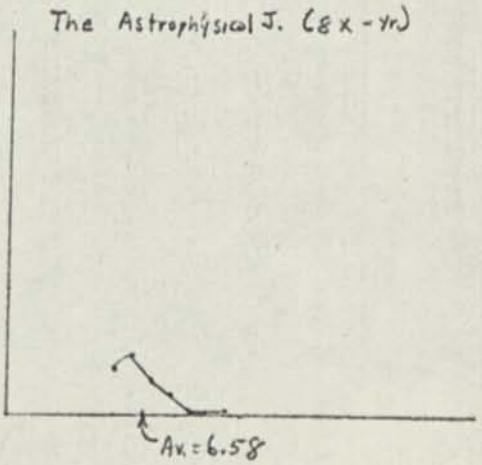
J. of Phys. Chem. (mo.)



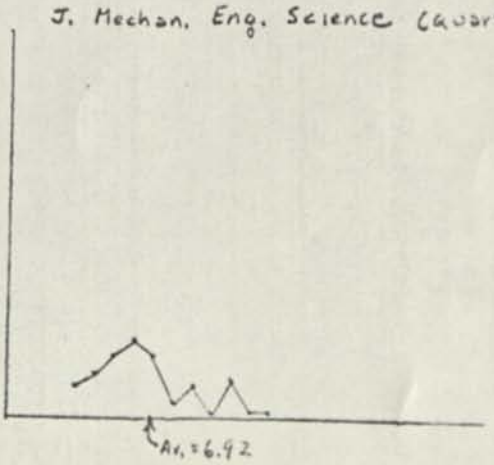
J. of Econ. Entomol. (bi-mo.)



The Astrophysical J. (8 x - yr)



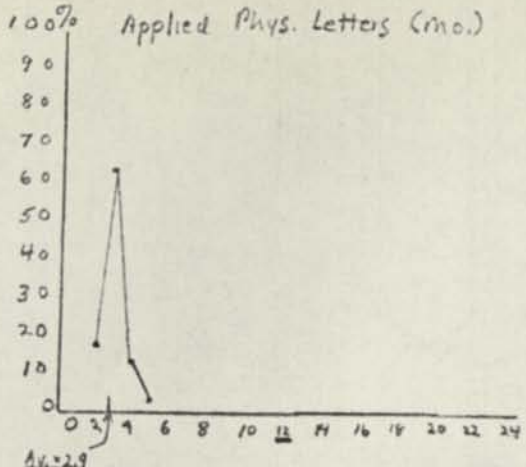
J. Mechan. Eng. Science (quarterly)



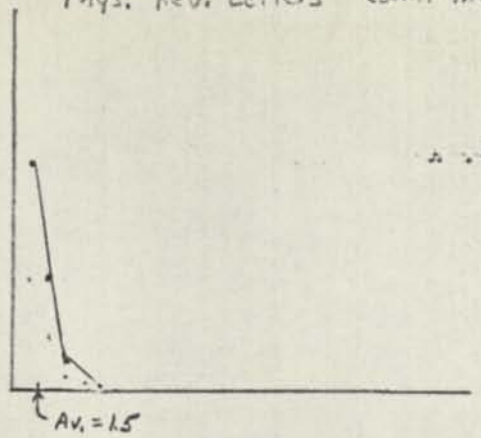
Percent of total articles with this time lag.

"Time lag" = delay from receipt of edited manuscript to date of publication of the Primary Publication. (Cover date)

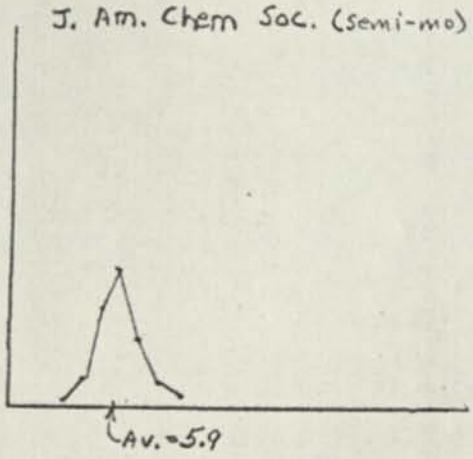
Sample Data in each chart is based on 50 articles appearing in most recent issue(s) of the primary publication.



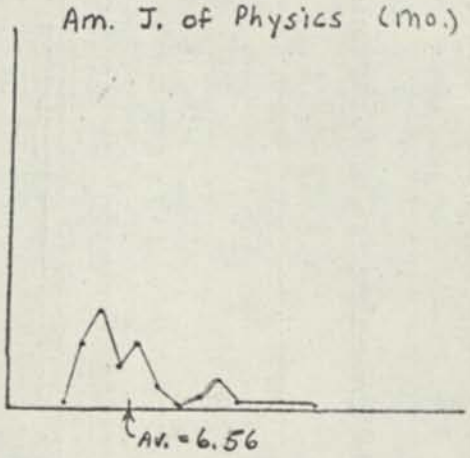
Applied Phys. Letters (mo.)



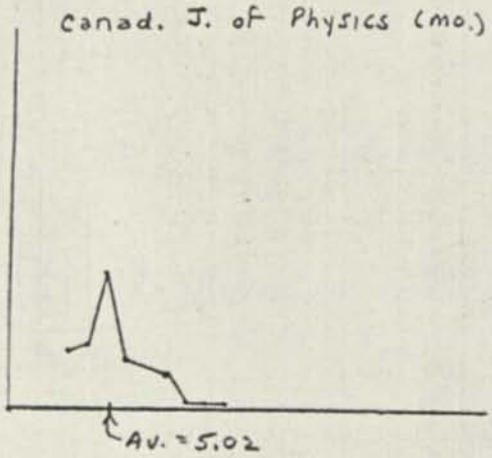
Phys. Rev. Letters (semi-mo)



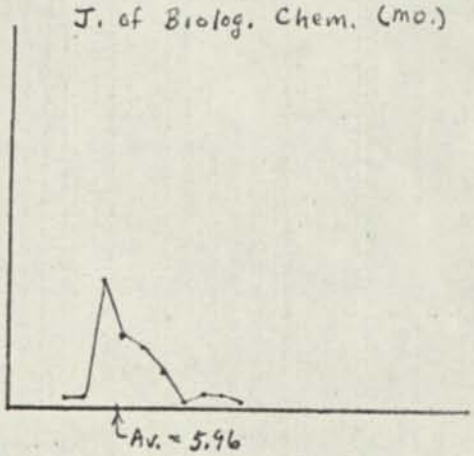
J. Am. Chem Soc. (semi-mo)



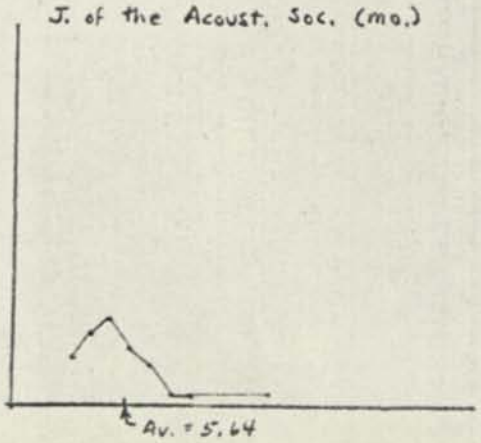
Am. J. of Physics (mo.)



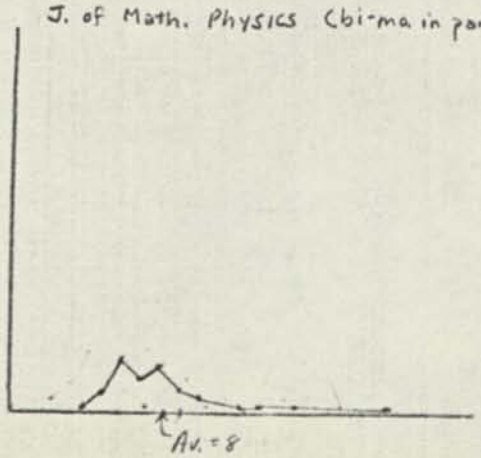
Canad. J. of Physics (mo.)



J. of Biolog. Chem. (mo.)



J. of the Acoust. Soc. (mo.)



J. of Math. Physics (bi-mo. in part)

Class Problem No. 5 TIME LAGS IN SECONDARY PUBLICATIONS
Librarianship 245 (Intro. to Documentation)
Charles P. Bourne, Instructor

Abstracting and indexing journals have often been referred to as "secondary" publications since they usually follow and report the prior publication of material in the primary journals. They may be used for current awareness reporting to keep a reader informed of work in his field, or they may be used for retrospective search tools. They need not be designed or used for both purposes. As a matter of fact, some of the abstract journals have become too large (e.g. Chemical Abstracts) to permit convenient cover-to-cover scanning by the reader.

In any case, the secondary publications serve an important role in these and other ways, and it is important that they report their material with a minimum time lag between primary publication and secondary reporting. Some critics claim that "...delays of a year or more are common for many of the secondary publications." If we are to use these secondary publications as everyday tools, we should be aware of their possible shortcomings and limitations. We should also be able to objectively examine such criticism as that stated above, in order to support or reject it on some rational basis. That is precisely what we shall do with this particular exercise, using some very simple sampling techniques.

Each student will examine 50 abstracts or index entries in the most recent issues of one of the secondary publications listed below. (A different publication for each student.) Taking the first full abstract or index entry on each even-numbered page, determine the time lag (to the nearest month) between the date of the original article or report, and the date of that issue of the secondary publication. Tally the frequency of occurrence of the 50 time lags, and the other requested data, in the spaces provided on the attached sheet. Plot the time lag data in the graph provided. The information from these summary notes will be assembled into final form by the Reporter assigned to this particular problem. The report submitted by the Reporter will then be reproduced and distributed to the class members.

ABSTRACT JOURNALS TO BE STUDIED

Abstracts of Photographic Science and Engineering Lit.
Abstracts of Soviet Medicine (Part A)
Abstracts of Soviet Medicine (Part B)
ACM Computing Review
Agricultural Index
API Technical Abstracts
Applied Science and Technology Index
Applied Mechanics Review
ASM Review of Metal Literature
Battelle Tech. Review (Abstracts Section)
Bibliography of Agriculture
Biological Abstracts
Business Periodicals Index
Cancer
Cardiovascular Diseases
Corrosion Abstracts
Crerar Metals Abstracts
Economic Abstracts
Electrical Eng. Abstracts
Engineering Index
Environmental Effects on Materials and Equipment
Gas Abstracts
Geophysical Abstracts
Historical Abstracts
IRE Trans. on Electronic Computers (Abstracts Section)
International Abstracts in Operations Research
International Aerospace Abstracts
J. Amer. Ceramic Soc. (Ceramic Abstracts Section)
J. Acoustical Soc. Amer. (Reference Section)
Meteorological and Geoastrophysical Abstracts
NASA Technical Publications Announcements
National Petroleum Bibliography
Nuclear Science Abstracts
Public Health Eng. Abstracts
Readers' Guide to Periodical Lit.
Scientific and Technical Aerospace Reports (STAR)
Semiconductor Products
Technical Abstract Bull. (ASTIA)
Tobacco Abstracts
U. S. Govt. Research Reports

Student's Name _____

Name of Secondary Publication Studied: _____

Publisher: _____

Frequency of Publication _____

Dates of secondary publication issues that included the 50 chosen abstracts or index entries: _____ to _____

No. of abstracts or index entries with this time lag (in months):

1 2 3 4 5 6 7 8 9 10 11 12

13 14 15 16 17 18 19 20 21 22 23 24 >2 yrs.

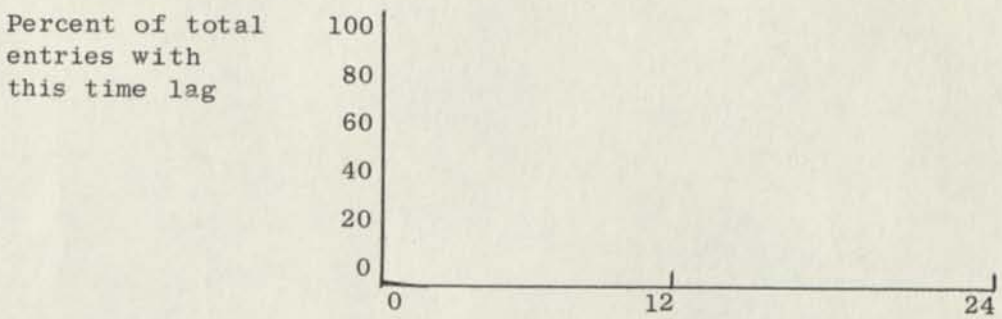
Percent of total (50) abstracts or index entries with this time lag (in months):

1 2 3 4 5 6 7 8 9 10 11 12

13 14 15 16 17 18 19 20 21 22 23 24 >2 yrs.

Average time lag:

Plot your percentage data and average on the chart below:



Approx. amount of time spent on this exercise: _____.

6 May 1963

Class Problem No. 5 TIME LAGS IN SECONDARY PUBLICATIONS
Librarianship 245 (Intro. to Documentation)
Class Reporter: Arnold Miller

1. Introduction

This project studied the time lag between the appearance of articles in primary journals and their reporting in the secondary journals.

Abstracting and indexing journals presently play an important role in major fields of study and research. To avoid wasteful duplication of effort, and for other reasons, it is important that the secondary journals publish their material with a minimum time lag between primary publication and secondary publication. The problem presented to the individual student in this exercise was to determine time lags in an assigned secondary journal, using simple sampling techniques. Each student examined the time lags for the 50 most recent abstracts or index entries in his assigned secondary journal. The student then tallied the total number of articles with various delays. The average time lag and the percentage of articles found in each time lag group was then computed and plotted.

This report is a consolidation of the individual student reports plus a conclusion reached by the reporter.

2. General Conclusions

The time lag in many areas is long. I believe that an examination of the time lags in the four important journals below will show that the best possible work is not being done in certain critical fields:

	Time Lag (in Months)
Economic Abstracts	3.58
International Aerospace Abstracts	3.82
Battelle Technical Review (Abs. Sect)	4.00
Nuclear Science Abstracts	6.36

It appears that the time lag in the field of nuclear science is not only undesirable but also susceptible to improvement when compared to the three other journals. Certainly the need for rapid publication of abstracts in the nuclear science field is equal to or greater than the need in the field of economics. It appears that an inquiry should be made as to the reason for the disproportionate lag in nuclear science.

The longest time lag appears in Sociological Abstracts. The very long average of 28.3 months appears to have several causes. First, the journal is a relatively new one and may now be engaged in retrospective abstracting of the important articles of recent years. If this is so, then this factor may tend to disappear and therby reduce the time lag average. It is also generally true that in the fields of social science the pressure and urgent need for rapid abstracting is not so severe as in the natural and physical sciences. It may also be true that monetary support for sociology is not as great as it is in natural sciences. This might also be a factor in the time lag.

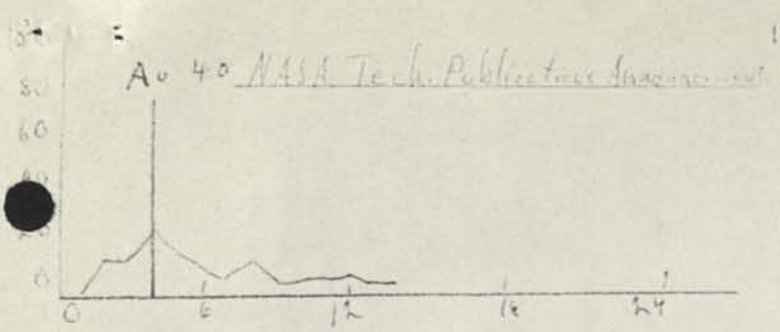
In estimating the effect of the time lags one should remember that the lags noted in this report are in addition to the lags between submission of the articles and their appearance in primary journals. The total lag will be considerable in many instances. There certainly exists a possibility to make unnecessary duplication of research work during the period of lag.

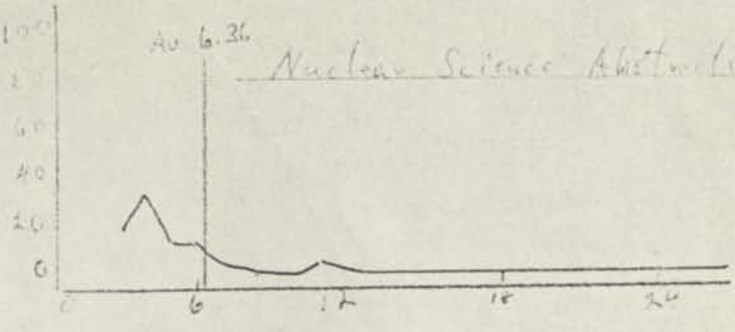
It appears that an integrated survey of time lag in publications in all the major fields might be profitably undertaken.

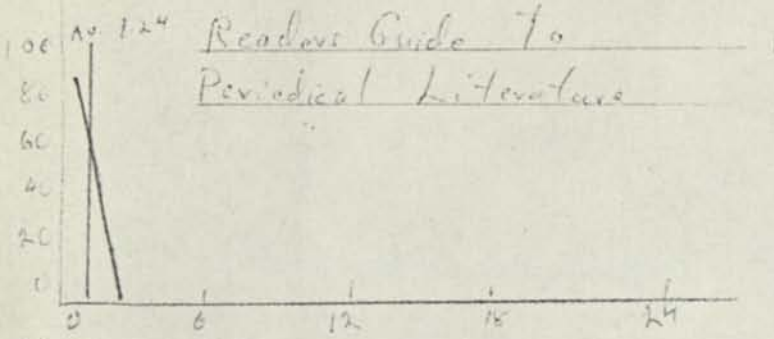
3. Consolidated statistical report covering time lag in 16 secondary publications.

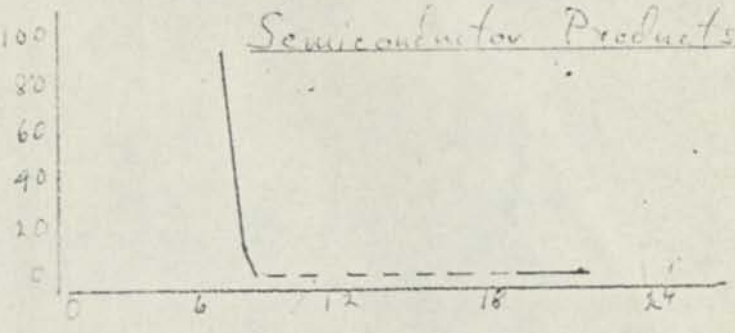
Average time lag all publications: 5.6 months.

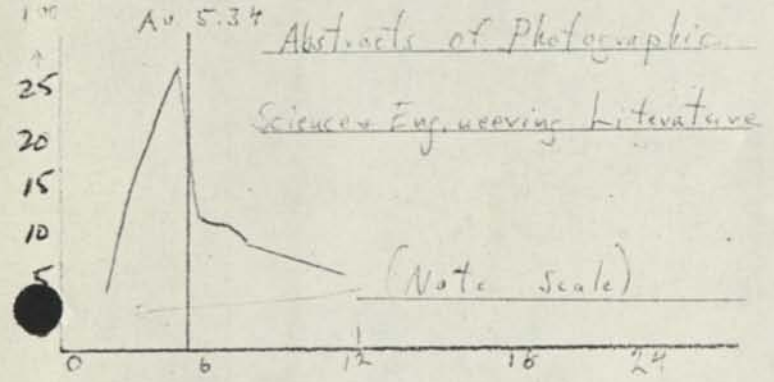
Time Lag (mos.)	No of abstracts or Index Entries with this Lag	Percent of Total Abstract or Index Entries with this Lag
1	41	5.1
2	58	7.2
3	116	14.5
4	106	13.2
5	126	15.7
6	90	11.2
7	86	10.7
8	32	4.0
9	12	1.5
10	16	2.0
11	14	1.7
12	10	1.2
13	5	0.6
14	10	1.3
15	6	0.8
16	4	0.5
17	0	0
18	1	0.1
19	7	0.9
20	6	0.8
21	6	0.8
22	3	0.3
23	4	0.5
24	5	0.6
over 24	51	4.0

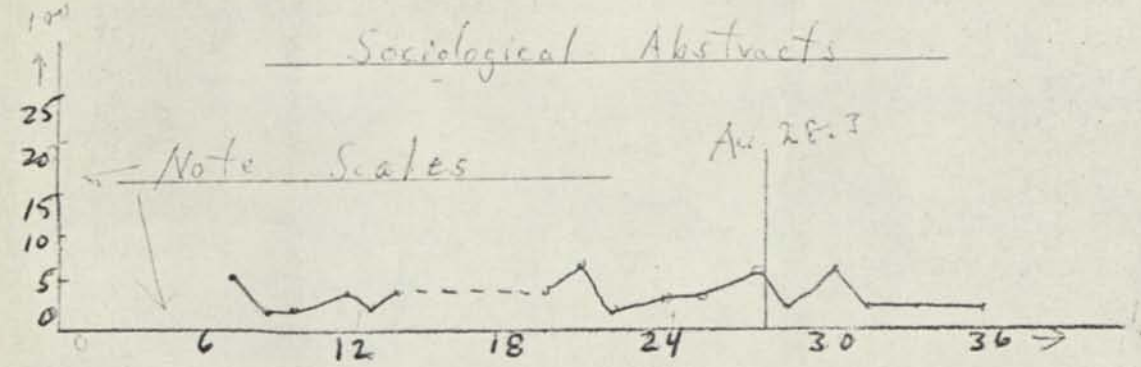

Av 40
NASA Tech. Publications Announcement
100
80
60
40
20
0
0
6
12
18
24

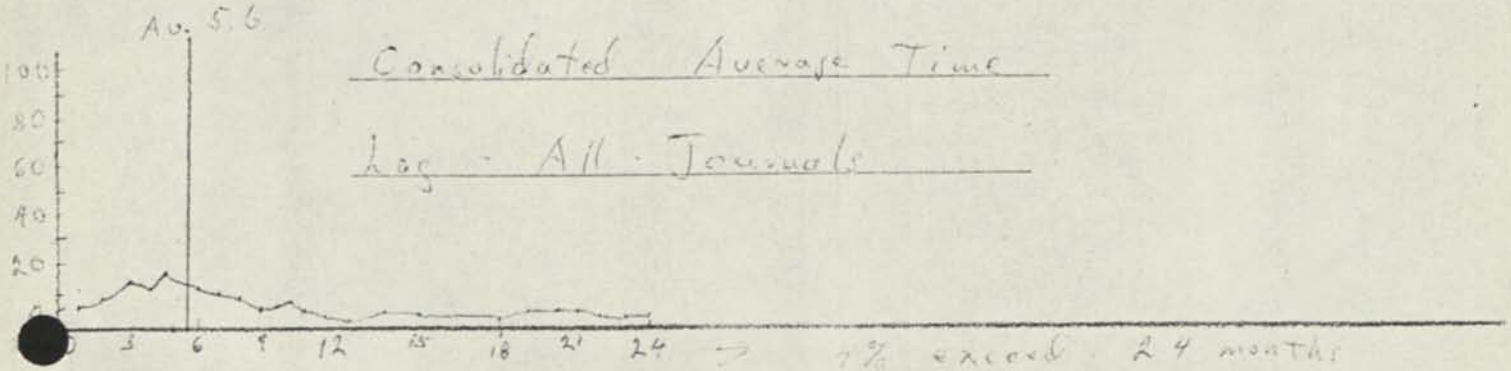

Av 6.36
Nuclear Science Abstracts
100
80
60
40
20
0
0
6
12
18
24


Av 1.24
Readers Guide To
Periodical Literature
100
80
60
40
20
0
0
6
12
18
24

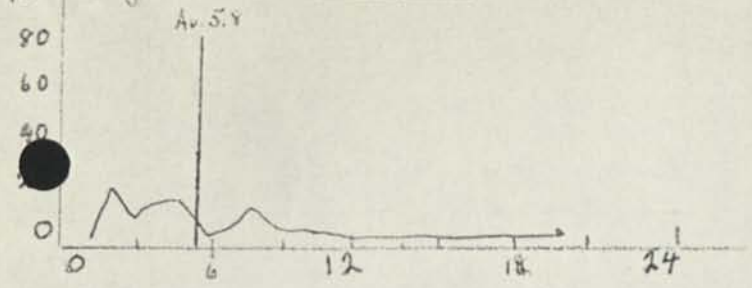

Semiconductor Products
100
80
60
40
20
0
0
6
12
18
24


Av 5.34
Abstracts of Photographic
Sciences Engineering Literature
(Note Scale)
25
20
15
10
5
0
6
12
18
24

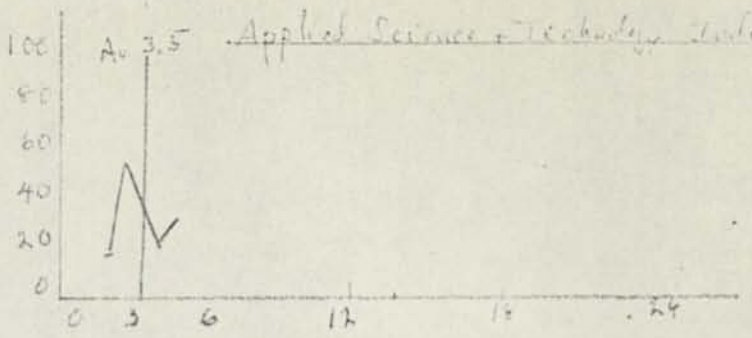

Sociological Abstracts
Note Scales
Av 28.3
25
20
15
10
5
0
6
12
18
24
30
36


Av. 5.6
Consolidated Average Time
Lag - All Journals
100
80
60
40
20
0
3
6
9
12
15
18
21
24
75% exceed 24 months

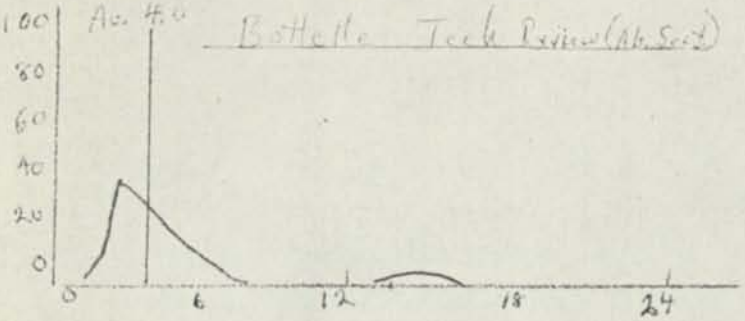
Agricultural Index



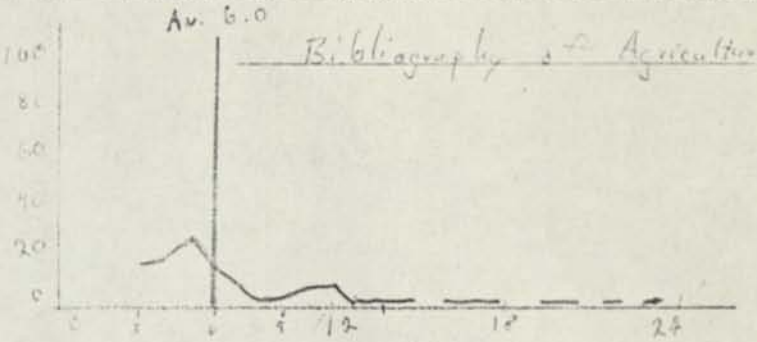
Applied Science & Technology Index



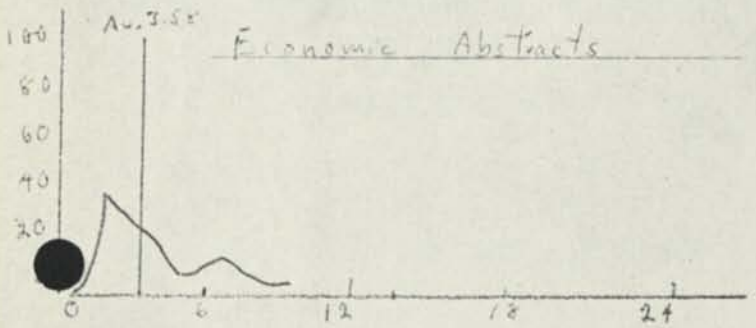
Bothelle Tech Review (Ab. Sect)



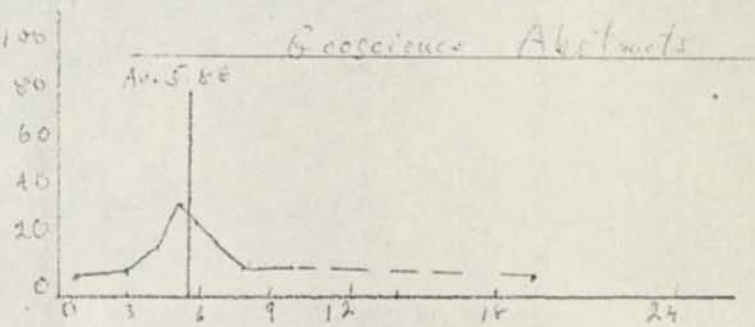
Bibliography of Agriculture



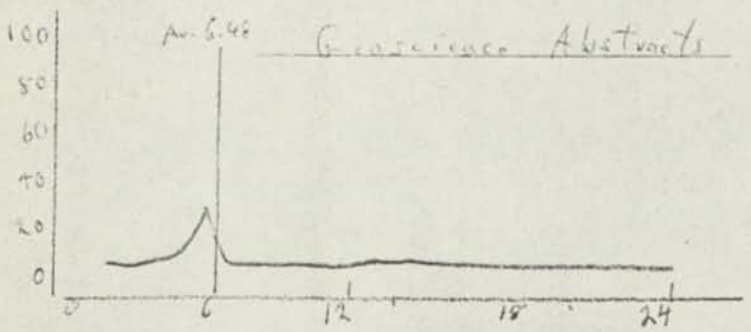
Economic Abstracts



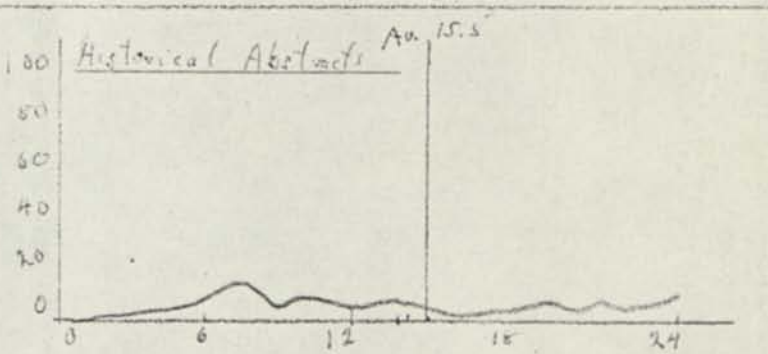
Geoscience Abstracts



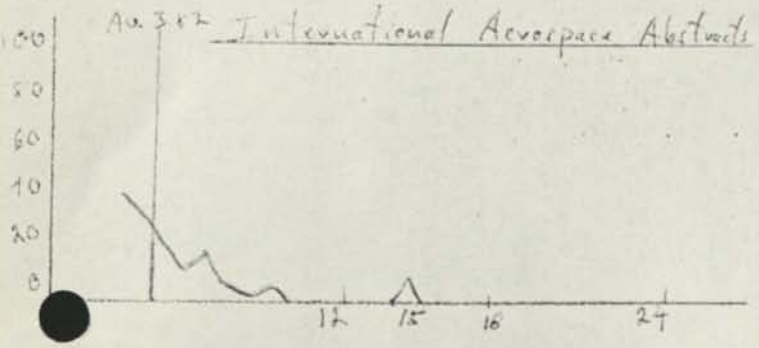
Geoscience Abstracts



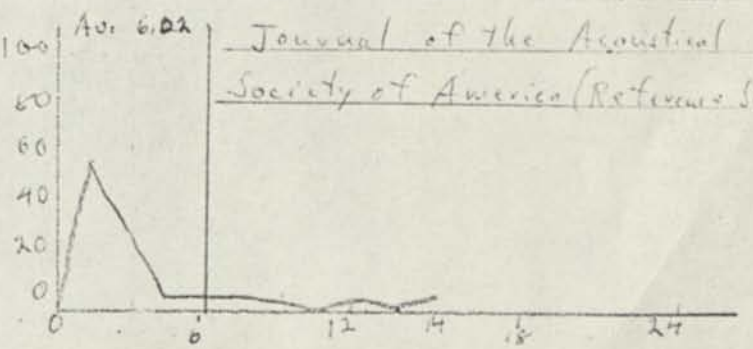
Historical Abstracts



International Aerospace Abstracts



Journal of the Acoustical Society of America (Reference Sect)



Class Problem No. 6 CODES AND NOTATIONS
Librarianship 245 (Intro. to Documentation)
Charles P. Bourne, Instructor
SUPERIMPOSED AND PRIME NUMBER CODES

From the class work and reading assignment, the student should be familiar with the methods for coding and searching with superimposed and prime number codes. This exercise provides some drill in the use of these codes. The following code table will be used for the subsequent questions.

Descriptor	Prime Number Code	Code Pattern for Superimposed Coding
Activation	2	010001000
Admittance	3	001000010
Alpha particle	5	000010001
Amplification	7	110000000
Angular frequency	11	000100010
Anode	13	101000000
Antennas	17	000110000
Apertures	19	000010010
Approximations	23	100010000
Arc	29	000000101
Argon	31	000001001
Atom	37	000010100
Autosyn	41	001000100
Avalanche	43	001000001
Bandwidth	47	000001010
Barrier	53	001010000
Battery	59	000001100
Beam	61	001100000
Bessel Function	67	100000100
Bevatron	71	010000001

1. What would the composite indexing notations be for the following sets of descriptors?

Descriptor Set	Prime No. Index Notation	Superimposed Code Notation
Activation, Anode, Arc,	2·13·29	11100 1101
Antenna, Apertures,	17·19	0001 1 0010
Bevatron, Beam, Atom, Barrier	37·53·61·71	0 1111 0101
Amplification, Avalanche, Barrier	7·43·53	1110 1 0100
Argon, Arc, Beam, Amplification, Atom	7·29·31·37·61	1111 1111

2. Which of the following documents are concerned with Argon?

Document No.	Prime No. Index Notation
(1)	54,653
2	56,287
3	4,823
4	8,700,553
5	2,485

etc.

31 | 54653 → 1763 (31, 236, 217, 195, 186, 93, 93)

1147 | 54653 → 47 (4588, 8773)

3. Which of the above documents are concerned with both argon and atom?

not 5 or 3 or 2 or 1

31, 37, 517, 93, 1147

1147 | 56287 → 49 (4588, 10407)

4. Which of the following documents are concerned with Bevatrons?

Document No.	Superimposed Code Notation (Composite)
(1)	111111011 ✓
2	101110010
3	001011101
4	001011111
(5)	111100101 ✓

5. Which of the above documents are concerned with both antennas and apertures?

1 & 5

NOTATION

1. For the following subject headings, provide the particular notation that has been designated by the indexing or classification schemes.

LC Subject Heading*	~~Dewey~~	Library of Congress	~~UDC~~ ~~Bliss~~ ~~ASTIA coord. index~~
✓religion		BL-BX	
✓Electronic digital computers		QA 76.5; QA 374; QA 297; TP 151; TK 7888.3; HF 5548; QA 75	
✓horse-racing† **		SF 301; SF 321-359	
✓jet ~~air~~planes		TL 571; TL 573	
✓documentation		Z1001; Z695.1; Z695.92; T11; Z1008; Q101; Z699.5; Z1007	
~~ball point pens~~			
✓Pens; ~~Writing - materials, instruments~~		~~TS 12~~	

We have no books in library on anything but steel pens

2. Cite the longest example of notation that you can find for each of the schemes mentioned above.

Scheme	LC Subject heading*	Notation
~~Dewey~~		
LC	Transistors	TK 7872.T73 (+ cutter no. L76t) (Transistors) (Linvill, JG)
~~UDC~~		
~~Bliss~~		
~~ASTIA~~		

* If you aren't using special subject heading lists are you? or Sears?

† on this one, I cheated - I looked in the classification tables, + not the catalog!

** Jet planes - Motors

See Aeroplanes - Jet propulsion
Aeroplanes - Two jet engines } not included in call no.

You would have to see both and classification table to understand why each number was used.

Descriptor Set	Prime No. Index Notation	Superimposed Code Notation
Activation, Anode, Arc,	2×13×29 = 754	111001101
Antenna, Apertures,	17(19) = 323	000110010
Bevatron, Beam, Atom, Barrier	71(41)(37)(53) = ~~8,493,091~~	011110101
Amplification, Avalanche, Barrier	(71)(43)(53) = 15,953	111010001
Argon, Arc, Beam, Amplification, Atom	(31)(29)(61) = 14,203,301 (17)(37)	111111101

01000 1000 101 101 → 111001101

00011 10010 → 000110010

2a. Which of the following documents are concerned with Argon?

Document No.	Prime No. Index Notation
(1)	54,653 only the
2	56,287 Notations for Doc #1, 4 can be divided by 31 with no remainder.
3	4,823
(4)	8,700,553
5	2,485

2b. ~~3.~~ Which of the above documents are concerned with both argon and atom?

None. i.e. None of the Composite Notations could be divided by 31×37 with no remainder.

3a. ~~4.~~ Which of the following documents are concerned with Bevatrons?

Document No.	Superimposed Code Notation (Composite)
(1)	111111011
2	101110010 Only the notations for Doc #1 & 5 include the pattern
3	001011101 010000001
4	001011111
(5)	111100101

3b. ~~5.~~ Which of the above documents are concerned with both antennas and apertures?

Only Doc # 1 and 2 included the composite pattern 000110010

NOTATION

1. For the following subject headings, provide the particular notation that has been designated by the indexing or classification schemes.

Subject Heading	Dewey	Library of Congress	UDC	Bliss	ASTIA coord. index
religion (681.14 681.142 30)	200	BL48 | BL1-2790 BL48-50 | BL-BZ BL-BX | BL	2 | 2.0. | 21	P | PA	
digital computers (681.42 | 510.78 | 510.783 | 681.14)		QA76 | QA75 | Z699.3 | QC54 4C6	681.142 | 681.14 | 681.3		
horse racing					
jet airplanes					
documentation					
ball point pens					

2. Cite the longest example of notation that you can find for each of the schemes mentioned above.

Scheme	Subject heading	Notation
Dewey		
LC		
UDC		
Bliss		
ASTIA	all subject headings have the same length of notation, — digits	

LC No. - accepted

religion	BL 48 or BL48-50 or BL-BX
digital computers	QA 75
horse racing	SF 334 or SF 321-359
jet airplanes	TL 670, N TL 709 or TL658.J4 or TL 670-688
documentation	Z 699
ball point pens	Z 45 or TS 1262 or TS 1263 or HD 9999. P37

Bliss No. - accepted

religion	P
digital computers	TJK
horse racing	HLT or QKS
jet airplanes	BTD or TNW
documentation	ZM or ZAD-H
ball point pens	TY

longest LC no. excluding the date = 12 HS1510.R3W752
Report to Grant Council, N.Y. Pocahontas
longest Bliss no. = 5 NNNYV, NQWWS, NNMYM
Vicksburg, Miss.; Sitka, Alaska; and Mobile, Ala.
respectively

Universal Decimal classification. Best notations

	2 Religion	681.14 Digital computers	798.4 Horse Racing	629.13.035 Jet airplanes	002 Documentation	686.863.7 Ball point pens.
Simeon						
Braaten	21 Natural Theology +	none +	798. equestrian spot animal racing OK	629.13 aircraft OK		686.8 stationery +
Clark						
Dobler						
Dye			636.12 race horses, breeds +			686.863 OK Pens, pencils
Fair		681.142 computers OK				
Horn	no notation +					
Leo		none +	798 OK	629.13 OK		686.863 OK
Miller		681.3 Apparatus in blown glass +				
Parsons		681.142 OK				
Reilly				629.135.2 Aerodynes. Aeroplanes +		
Schwartz						686.7 mirrors +
Shain						
Siebert		681.3 +				
Silvera						686.7 +
~~Siemon~~	no notation +	~~none~~ +	~~none~~ +	~~none~~ +	~~none~~ +	~~none~~ +
Skelley						
Stephens						686.7 +
Tarcy						
Stys						

OK variant notation acceptable
+ notation in the heading more specific

Longest notation:

Reilly : Dewey : 331.818.588.709.42
~~Adams~~ ~~U.D.C. 621.396.62~~
Miller : U.D.C. : 621.365 : 643.33

Prepared by Stephen D. Tarcy

Dewey. Best notations

	510.783 digital computers	629.133.349 jet airplanes	688 manufacture of small articles	uniform answers:
Braaten				religion 200
P.F Clark				horseracing 798.4
Dobler				documentation 010
Dye	681.14 OK Business machines		651.71 OK Stationery supplies	
Fair	681.14 ✓ OK			
Horn				
Leo				
Miller				
Parsons	681.42 leased X			
Reilly				
Schwartz				
Shain	629.134 353 jets	629.134.353 X gas turbine & jet engines	679 X mfr of misc products	
Siebert				
Silvera				
Siemon				
Skelley		629.134.353 X		
Stephens				
Tarcey		629.134.353		
Stys		387.733 49 X air transport. airplanes.		

OK variation acceptable, no specific class exists.
X choice in the heading more specific

Prepared by Stephen J. Tarcey

Class Problem No. 8 THE USE OF A COMPUTER TO PREPARE INDEXES, CATALOG CARDS, AND ACCESSION LISTS

Librarianship 245 (Intro. to Documentation)

Charles P. Bourne, Instructor

At one time or another in the last few years, some libraries have found a way to use a tab card or computer equipment to prepare accession lists, or catalog cards, or printed indexes (usually one application per library attempt). More recently, several libraries working independently have found ways to use this equipment to do all of these jobs for the same library, utilizing the same input information. With enough planning and attention to system design, any general bibliographic information can be punched into cards (with some constraints on the formatting and arrangement of information) which can then be used as the basis for a complete package of library routines.

As an exercise in the use of punched card equipment, and an illustration of such a multi-purpose computer program for library applications--the students will keypunch a collection of citations (90 abstracts in the Literature Notes section of the April 1963 American Documentation) which will then be processed using the computer and programs of the U.C. Radiation Lab. The results of this effort will be a KWIC index, catalog cards, and accession list for these 90 citations.

In order for the programs to work correctly, careful attention must be paid to the details of card punching. Specific instructions, tab listings of sample sets of cards, sample catalog cards, and a keypunch work sheet are given below. To save time at the keypunch facility, the assigned abstracts should be transcribed ahead of time onto the worksheets, and checked for form and accuracy with other students. (Remember, if you goof, your name will be printed out along with the bum citation!)

GENERAL INSTRUCTIONS FOR CARD PUNCHING

1. Keypunching Control Numbers

Columns 1 thru 7 = serial numbers. These numbers or letters should be duplicated (the same) in every punched card for a given citation. These columns can represent anything that you wish, e.g., type of list, subject category, etc. Do not place an 8 in column 4.

Columns 8, 9 = line number (punched card sequence). The first line of any citation should always be = 01. The remaining line numbers should be in ascending order, but need not be sequential, i.e., either 1, 2, 3, or 1, 3, 5, 31 is all right.

Some line numbers have special functions. These are:

Line 01 = will instruct the computer to skip a line before printing the next line on catalog cards only. It also indicates that this is the beginning of a new citation.

Line 25 = 2 lines will be skipped on the catalog cards only before printing the next line.

Line 31 = end of the list citation when doing a list. When catalog cards are being produced 2 lines will be skipped before printing the next line.

Line 41 = end of the list citation. When catalog cards are printed, no lines will be skipped.

Lines 51, 61, 71 = the same as line 25.

Line 91 = three lines will be skipped after the printing of the line on catalog cards only.

Columns 10, 11 on lst line of the citation only = number of catalog cards to be made for this citation. If no catalog cards are to be made, then place 2 zeros in these positions.

2. Keypunching the Citation Data

1. Place the information on the punched cards starting in column 10 (except on list card) in any desired format. Each punched card will represent a citation line of information.

2. Place the main entry on card 1 beginning in column 12. (This could be a call number, an accession or report number, a volume and abstract number, etc.) For this exercise, use the first personal author (or corporate author if there is no personal author) as the main entry.

3. Only 46-47 characters per line should be typed, besides the 9 number positions.

4. If a line or lines is not to be put on the catalog cards but printed on the lists, place an equal sign, =, after the last character of the line in question. Therefore, an equal sign that should be printed on the catalog cards or the lists should not be positioned at the end of a line.

5. The computer has a limited number of special symbols. Try to use them, only. These symbols are: period, comma, slash, equal sign, dollar sign, plus, dash, apostrophe, open and closed parenthesis, and an asterisk.

6. Place a dollar sign, $, before and after the citation portion to be indexed. The portion may be title alone, title plus author*, abstract, depending on what type of index you need. The data to be indexed can appear on as many lines as desired. These lines must follow each other without a break in the sequence. Thus, the initial dollar sign, $, will always be in column 10. The final dollar sign, $, will follow the last character of the portion to be indexed. Never place any other data on the same line following the second $. Always begin a new line. For this exercise, enclose only the title in dollar signs.

*If authors are to be indexed, use their first and middle initials only, besides the surname.

05

SALTON, GERARD. 1961.

INFORMATION STORAGE AND RETRIEVAL.
SCIENTIFIC REPORT NO. ISR-1 ON CONTRACT
AF 19(604)8509, PROJ. 5632. AFCRL-62-77.
COMPUTATION LABORATORY, HARVARD UNIVERSITY,
CAMBRIDGE, MASS.
152P
AVAILABLE AS AD-274816 FROM THE OFFICE OF
TECHNICAL SERVICES, WASH. D.C., FOR $2.75.

(ABSTRACTED IN AMER. DOC.--JAN., 1963, P89)

05

SALTON, GERARD. 1961.

INFORMATION STORAGE AND RETRIEVAL.
SCIENTIFIC REPORT NO. ISR-1 ON CONTRACT
AF 19(604)8509, PROJ. 5632. AFCRL-62-77.
COMPUTATION LABORATORY, HARVARD UNIVERSITY,
CAMBRIDGE, MASS.
152P
AVAILABLE AS AD-274816 FROM THE OFFICE OF
TECHNICAL SERVICES, WASH. D.C., FOR $2.75.

(ABSTRACTED IN AMER. DOC.--JAN., 1963, P89)

05

05

SALTON, GERARD. 1961.

INFORMATION STORAGE AND RETRIEVAL.
SCIENTIFIC REPORT NO. ISR-1 ON CONTRACT
AF 19(604)8509, PROJ. 5632. AFCRL-62-77.
COMPUTATION LABORATORY, HARVARD UNIVERSITY,
CAMBRIDGE, MASS.
152P
AVAILABLE AS AD-274816 FROM THE OFFICE OF
TECHNICAL SERVICES, WASH. D.C., FOR $2.75.

(ABSTRACTED IN AMER. DOC.--JAN., 1963, P89)

05

CARD LISTING (80-80)

```
10317920105ASD-TDR-63-144(PT.1)
103179202AMERICAN OIL CO. RESEARCH AND DEVELOPMENT
103179203DEPT, WHITING, IND
103179204$RADIATION CHEMISTRY OF TRIPLE BOND
103179205COMPOUNDS. PART 1. ACETONITRILE AND
103179206BUTYNE-2, LIQUID PHASES$
103179207T D NEVITT, R E RONDEAU, L A HARRAH ET AL
103179231FEBRUARY 1963  26P
103179251   (CONTRACT AF-33(616)-8247)
103179252R.1C.630304-MC
10217000105NRDL-TR-606
102170002NAVAL RADIOLOGICAL DEFENSE LAB, SAN FRANCISCO
102170003$EFFECTS OF DIVIDED DOSES OF X-RAY ON
102170004MORTALITY AND HEMATOLOGY OF SMALL AND LARGE
102170005DOMESTIC ANIMALS$
102170006G F LEONG, W G WISECUP AND J W GRISHAM
102170031JANUARY 7, 1963  22P
102170032R.2C.630225-MC
10316420105AD-286653
103164202QUARTERMASTER RESEARCH AND ENGINEERING COM-
103164203MAND, NATICK, MASS
103164204$ORGANOMETALLIC RESEARCH USING GROUP IV
103164205ELEMENTS$
103164206M C HENRY, J G NOLTES, W DAVIDSON ET AL
103164231JUNE 1962  8P
103164232   (PRESENTED AT THE 1962 ARMY SCIENCE CON-
103164233FERENCE HELD AT U S MILITARY ACADEMY, WEST
103164251POINT, N Y, JUNE 20-22, 1962)
103164252R.1C.630225-MC
```

period no. — space — tells the no. of catalog cards to print

```
123556701100SBORN, A. D.
123556702$1962. THE INDEXING OF SCIENTIFIC LITERATURE.
123556703SIXTY MILLION PAGES A YEAR. THE MASSIVE PROBLEM
123556704OF SCIENCE INDEXING AND ABSTRACTING.$
123556731AUSTRALIAN J. SCI., 24 (11)  P429-437
123556732  ABSTRACTED IN AMER. DOC., JAN. 1963  P 89
```

"31" counts 2 blanks for the printing of the next line of the printing catalog cards

```
123657801105SALTON, GERARD.  1961.
123657802$INFORMATION STORAGE AND RETRIEVAL.$
123657803SCIENTIFIC REPORT NO. ISR-1 ON CONTRACT
123657804AF 19(604)8509, PROJ. 5632.  AFCRL-62-77.
123657805COMPUTATION LABORATORY, HARVARD UNIVERSITY,
123657806CAMBRIDGE, MASS.
123657807  152P
123657808   AVAILABLE AS AD-274816 FROM THE OFFICE OF
123657831   TECHNICAL SERVICES, WASH. D.C., FOR $2.75.
123657835     (ABSTRACTED IN AMER. DOC.--JAN., 1963, P89
```

6

5

4

3

2

EXAMPLE

cols. 1 5 6 7 8 9 10 55

4/63-	02	01	01	NAVAL ORDNANCE LAB
			02	$ PIONEERING IN MACHINE LITERATURE SEARCHING
			03	AND RETRIEVAL.$
				CYNAYWEPS REPT. NO. 7388; AD-255 291
				31 1961. 22pp. INCL. TABLES. OTS PRICE $2.60
			32	PUNCHED BY HAD NUFF

4/63-		01	01
			$

DO NOT GO BEYOUD COL. 55

NOTE: 1. The number 4/63- should be punched into the first 5 cols. of every card.

2. Cols. 6-7 represent the AD citation number, & should be punched with each card used for that citation.

3. Cols. 8-9 are for sequence numbers for each citation; they must be ascending but not necessarily sequential.

4. The first card for each citation should have a 01 in cols. 8-9 and cols. 10-11.

5. The title should be typed in $ signs, with nothing else on these cards.

6. The last card of the citation should have 31 punched in cols. 8-9. This card should be followed by another listing who prepared it.

PUNCHED CARD FORMAT FOR CATALOG CARDS, AN ACCESSION LIST, AND A KEYWORD-PLUS-TITLE INDEX

KEYPUNCHING CONTROL NUMBERS

Columns 1 thru 7 = serial numbers. These numbers or letters should be duplicated (the same) in every punched card for a given citation. These columns can represent anything that you wish, e.g., type of list, subject category, etc. Do not place an 8 in column 4.

Columns 8, 9 = line number (punched card sequence). The first line of any citation should always be = 01. The remaining line numbers should be in ascending order, but need not be sequential, i.e., either 1, 2, 3 or 1, 3, 5, 31 is al right.

Some line numbers have special functions. These are:

Line 01 = will instruct the computer to skip a line before printing the next line on catalog cards only. It also indicates that this is the beginning of a new citation.

Line 25 = 2 lines will be skipped on the catalog cards only before printing the next line.

Line 31 = end of the list citation when doing a list. When catalog cards are being produced 2 lines will be skipped before printing the next line.

Line 41 = end of the list citation. When catalog cards are printed, no lines will be skipped.

Lines 51, 61, 71 = the same as line 25.

Line 91 = three lines will be skipped after the printing of the line on catalog cards only.

Columns 10, 11 on <u>1st line</u> of the citation only = number of catalog cards to be made for this citation. If no catalog cards are to be made, then place 2 zeros in these positions.

KEYPUNCHING THE CITATION DATA

1. Place the information on the punched cards starting in column 10 (except on 1st card) in any desired format. Each punched card will represent a citation line of information.

2. Place a control number on card 1 beginning in column 12. This number can be a call number, an accession or report number, a volume and abstract number, etc. This number (less than 12 characters long) should be the only data on card 1, if the number is to be used for the reference number in the accession's list.

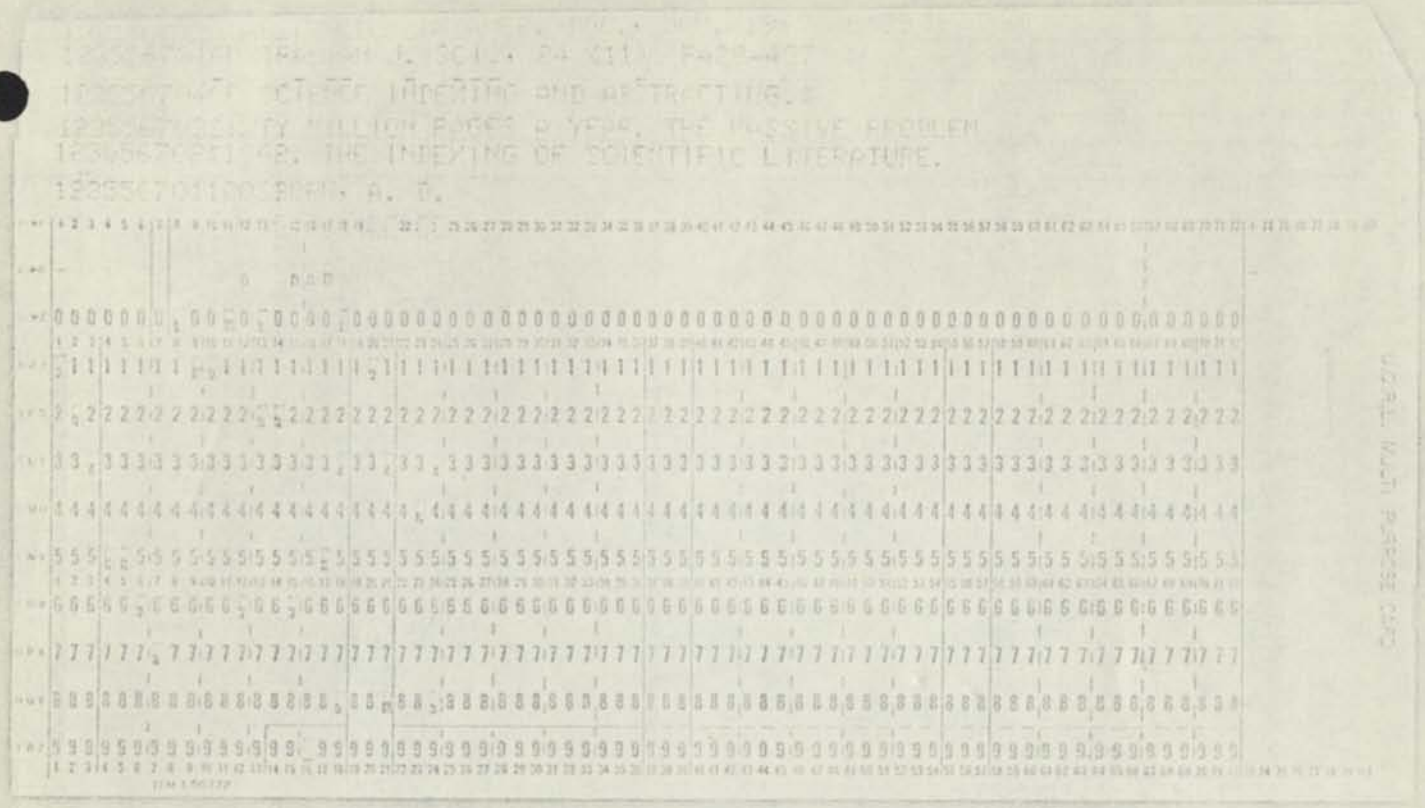
3. Only 46 - 47 characters per line should be typed, besides the 9 number positions.

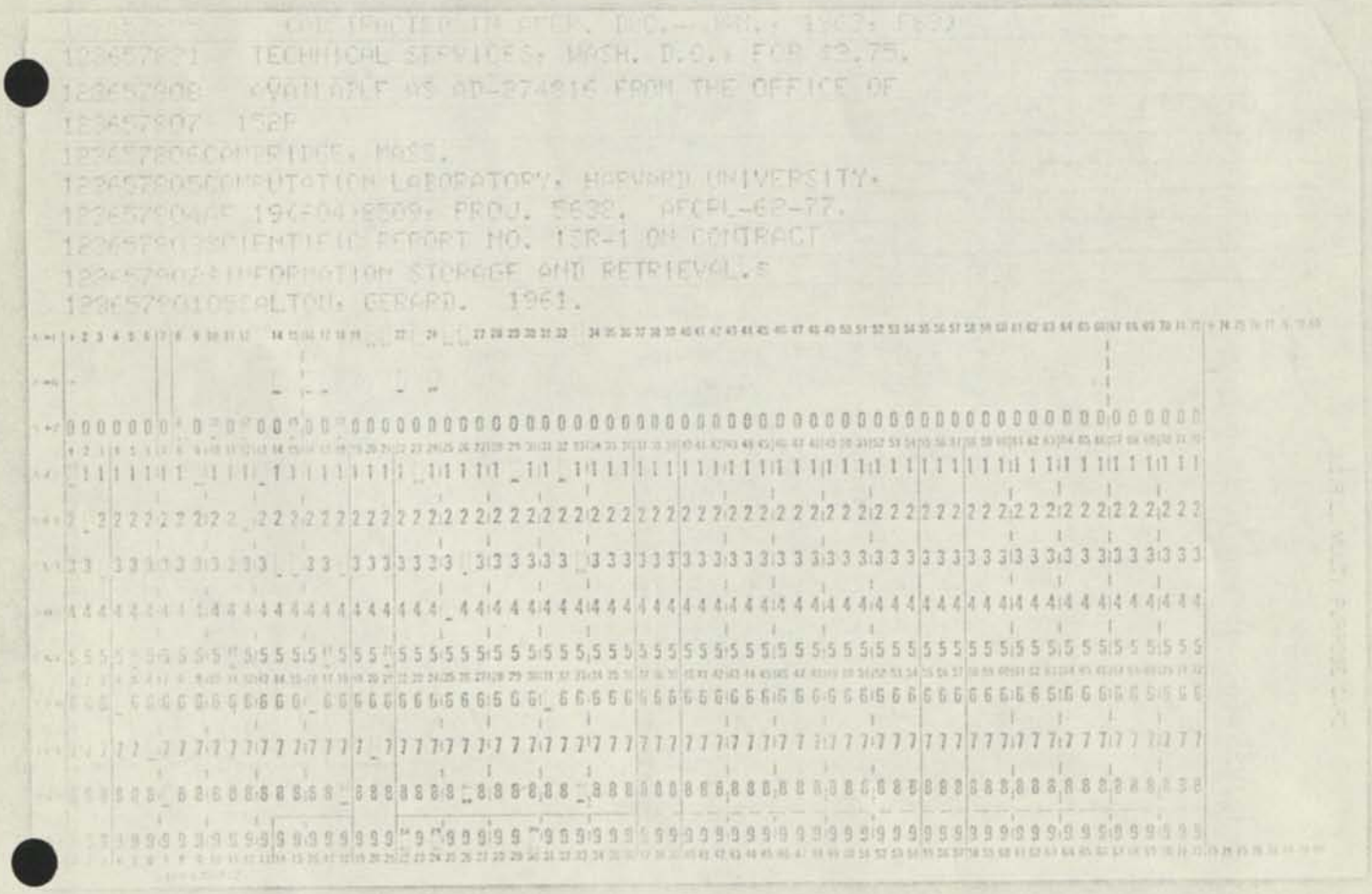
4. If a line or lines is not to be put on the catalog cards but printed on the lists, place an equal sign, =, after the last character of the line in question. Therefore, an equal sign that should be printed on the catalog cards or the lists should not be positioned at the end of a line.

5. The computer has a limited number of special symbols. Try to use them, only. These symbols are: period, comma, slash, equal sign, dollar sign, plus, dash, apostrophe, open and closed parenthesis, and an asterisk.

6. Place a dollar sign, $, before and after the citation portion to be indexed. The portion may be title alone, title plus author*, abstract, depending on what type of index you need. The data to be indexed can appear on as many lines as desired. These lines must follow each other without a break in the sequence. Thus, the initial dollar sign, $, will always be in column 10. The final dollar sign, $, will follow the last character of the portion to be indexed. Never place any other data on the same line following the second $. Always begin a new line.

*If authors are to be indexed, use their first and middle initials only, besides the surname.





UNIVERSITY OF CALIFORNIA

LAWRENCE RADIATION LABORATORY
P. O. BOX 808
LIVERMORE, CALIFORNIA

March 20, 1963

Mr. Charles P. Bourne
Research Engineer
General Systems Department
Stanford Research Institute
Menlo Park, California

Dear Charles:

Enclosed are key punching instructions. As you will note, the cards can be used for the production of catalog cards along with the list of citations and index. This multipurpose use may explain for the elaborate procedures.

The control number field is large enough, so that your students can learn about sorting the cards.

If this information is not adequate, please call me. I'll probably see you at the IBM workshop next week.

Sincerely yours,

Jim

James H. Kennedy
Documentation Specialist

JHK:aj
Enc.

reference number (hand-generated)

ASD-TDR-63-144(PT.1)
AMERICAN OIL CO. RESEARCH AND DEVELOPMENT DEPT, WHITING, IND
RADIATION CHEMISTRY OF TRIPLE BOND COMPOUNDS. PART 1. ACETONITRILE AND BUTYNE-2, LIQUID PHASE
T D NEVITT, R E RONDEAU, L A HARRAH ET AL
FEBRUARY 1963 26P

NRDL-TR-606
NAVAL RADIOLOGICAL DEFENSE LAB, SAN FRANCISCO
EFFECTS OF DIVIDED DOSES OF X-RAY ON MORTALITY AND HEMATOLOGY OF SMALL AND LARGE DOMESTIC ANIMALS
G F LEONG, W G WISECUP AND J W GRISHAM
JANUARY 7, 1963 22P

SWC-TDR-62-95
AIR FORCE SPECIAL WEAPONS CENTER, KIRTLAND AFB, N MEX
CLASSIFICATION OF CIRCULATING LEUKOCYTES IN THE NORMAL MEXICAN BURRO
R E ENGEL, S CARTWRIGHT AND F A SPURRELL
JANUARY 1963 12P

AD-286653
QUARTERMASTER RESEARCH AND ENGINEERING COMMAND, NATICK, MASS
ORGANOMETALLIC RESEARCH USING GROUP IV ELEMENTS
M C HENRY, J G NOLTES, W DAVIDSON ET AL
JUNE 1962 8P

computer generated #

UCRL-12000 VOL 5 NO 5

001 505

ASD-TDR-63-144(PT.I)
AMERICAN OIL CO. RESEARCH AND DEVELOPMENT DEPT, WHITING, IND
RADIATION CHEMISTRY OF TRIPLE BOND COMPOUNDS. PART 1. ACETONITRILE AND BUTYNE-2, LIQUID PHASE
T D NEVITT, R E RONDEAU, L A HARRAH ET AL
FEBRUARY 1963 26P

002 505

NRDL-TR-606
NAVAL RADIOLOGICAL DEFENSE LAB, SAN FRANCISCO
EFFECTS OF DIVIDED DOSES OF X-RAY ON MORTALITY AND HEMATOLOGY OF SMALL AND LARGE DOMESTIC ANIMALS
G F LEONG, W G WISECUP AND J W GRISHAM
JANUARY 7, 1963 22P

003 505

SWC-TDR-62-95
AIR FORCE SPECIAL WEAPONS CENTER, KIRTLAND AFB, N MEX
CLASSIFICATION OF CIRCULATING LEUKOCYTES IN THE NORMAL MEXICAN BURRO
R E ENGEL, S CARTWRIGHT AND F A SPURRELL
JANUARY 1963 12P

004 505

AD-286653
QUARTERMASTER RESEARCH AND ENGINEERING COMMAND, NATICK, MASS
ORGANOMETALLIC RESEARCH USING GROUP IV ELEMENTS
M C HENRY, J G NOLTES, W DAVIDSON ET AL
JUNE 1962 8P

```
10317920105ASD-TDR-63-144(PT.1)
103179202AMERICAN OIL CO. RESEARCH AND DEVELOPMENT
103179203DEPT, WHITING, IND
103179204$RADIATION CHEMISTRY OF TRIPLE BOND
103179205COMPOUNDS. PART 1. ACETONITRILE AND
103179206BUTYNE-2, LIQUID PHASES$
103179207T D NEVITT, R E RONDEAU, L A HARRAH ET AL
103179231FEBRUARY 1963  26P
103179251   (CONTRACT AF-33(616)-8247)
103179252R.1C.630304-MC
10217000105NRDL-TR-606
102170002NAVAL RADIOLOGICAL DEFENSE LAB, SAN FRANCISCO
102170003$EFFECTS OF DIVIDED DOSES OF X-RAY ON
102170004MORTALITY AND HEMATOLOGY OF SMALL AND LARGE
102170005DOMESTIC ANIMALS$
102170006G F LEONG, W G WISECUP AND J W GRISHAM
102170031JANUARY 7, 1963  22P
102170032R.2C.630225-MC
10316420105AD-286653
103164202QUARTERMASTER RESEARCH AND ENGINEERING COM-
103164203MAND, NATICK, MASS
103164204$ORGANOMETALLIC RESEARCH USING GROUP IV
103164205ELEMENTS$
103164206M C HENRY, J G NOLTES, W DAVIDSON ET AL
103164231JUNE 1962  8P
103164232   (PRESENTED AT THE 1962 ARMY SCIENCE CON-
103164233FERENCE HELD AT U S MILITARY ACADEMY, WEST
103164251POINT, N Y, JUNE 20-22, 1962)
103164252R.1C.630225-MC
```

```
123556701100SBORN, A. D.
123556702$1962. THE INDEXING OF SCIENTIFIC LITERATURE.
123556703SIXTY MILLION PAGES A YEAR. THE MASSIVE PROBLEM
123556704OF SCIENCE INDEXING AND ABSTRACTING.$
123556731AUSTRALIAN J. SCI., 24 (11)  P429-437
123556732  ABSTRACTED IN AMER. DOC., JAN. 1963  P 89
12365780105SALTON, GERARD.  1961.
123657802$INFORMATION STORAGE AND RETRIEVAL.$
123657803SCIENTIFIC REPORT NO. ISR-1 ON CONTRACT
123657804AF 19(604)8509, PROJ. 5632.  AFCRL-62-77.
123657805COMPUTATION LABORATORY, HARVARD UNIVERSITY,
123657806CAMBRIDGE, MASS.
123657807  152P
123657808   AVAILABLE AS AD-274816 FROM THE OFFICE OF
123657831   TECHNICAL SERVICES, WASH. D.C., FOR $2.75.
123657835     (ABSTRACTED IN AMER. DOC.--JAN., 1963, P89)
```

UCRL-12000 VOL 5 NO 6

0010010

OSBORN, A. D.
1962. THE INDEXING OF SCIENTIFIC LITERATURE.
SIXTY MILLION PAGES A YEAR. THE MASSIVE PROBLEM
OF SCIENCE INDEXING AND ABSTRACTING.
AUSTRALIAN J. SCI., 24 (11) P429-437

0020010

SALTON, GERARD. 1961.
INFORMATION STORAGE AND RETRIEVAL.
SCIENTIFIC REPORT NO. ISR-1 ON CONTRACT
AF 19(604)8509, PROJ. 5632. AFCRL-62-77.
COMPUTATION LABORATORY, HARVARD UNIVERSITY,
CAMBRIDGE, MASS.
152P
AVAILABLE AS AD-274816 FROM THE OFFICE OF
TECHNICAL SERVICES, WASH. D.C., FOR $2.75.

OSBORN, A. D.
1962. THE INDEXING OF SCIENTIFIC LITERATURE.
SIXTY MILLION PAGES A YEAR. THE MASSIVE PROBLEM
OF SCIENCE INDEXING AND ABSTRACTING.
AUSTRALIAN J. SCI., 24 (11) P429-437

SALTON, GERARD. 1961.
INFORMATION STORAGE AND RETRIEVAL.
SCIENTIFIC REPORT NO. ISR-1 ON CONTRACT
AF 19(604)8509, PROJ. 5632. AFCRL-62-77.
COMPUTATION LABORATORY, HARVARD UNIVERSITY,
CAMBRIDGE, MASS.
152P
AVAILABLE AS AD-274816 FROM THE OFFICE OF
TECHNICAL SERVICES, WASH. D.C., FOR $2.75.

UNIVERSITY OF CALIFORNIA

LAWRENCE RADIATION LABORATORY
P. O. BOX 808
LIVERMORE, CALIFORNIA

April 3, 1963

Mr. Charles P. Bourne
Research Engineer
General Systems Department
Stanford Research Institute
Menlo Park, California

Dear Charles:

Included are the punched cards for the output samples that I sent you several weeks ago. I also keypunched two citations from the Am. Doc., January 1963, Page 89. The punched cards and the output are included.

I have included a sample subject authority list prepared by the computer. Included, also, is a partial listing of the input tape, so you can see what information the computer automatically provided on the output.

Sincerely,

Jim

James H. Kennedy
Documentation Specialist

JKH:aj
Enc.

CHARACTER	CARD CODE	BCD CODE						
		C	B	A	8	4	2	1
BLANK	No Punches	X						
. period	12-3-8		X	X	X		X	X
( ) closed parenthesis	12-4-8	X	X	X	X	X		
(Undefined Special Character)	12-5-8		X	X	X	X		X
(Undefined Special Character)	12-6-8		X	X	X	X	X	
(Group Mark-Special Character) (Note 1)	12-7-8	X	X	X	X	X	X	X
+ plus	12	X	X	X				
$ dollar sign	11-3-8	X	X		X		X	X
* asterick	11-4-8		X		X	X		
(Undefined Special Character)	11-5-8	X	X		X	X		X
(Undefined Special Character)	11-6-8	X	X		X	X	X	
Δ (Mode Change (Delta) Special)	11-7-8		X		X	X	X	X
- dash	11		X					
/ slash	0-1	X		X				X
, comma	0-3-8	X		X	X		X	X
(open parenthesis	0-4-8			X	X	X		
(Word Separator-Special Char.)	0-5-8	X		X	X	X		X
(Undefined Special Character)	0-6-8	X		X	X	X	X	
Tape Segment Mark-Special Char.	0-7-8			X	X	X	X	X
1401 Generated Special Character (Note 2)				X				
= equals	3-8				X		X	X
' apostrophe	4-8	X			X	X		
(Undefined Special Character)	5-8				X	X		X
(Undefined Special Character)	6-8				X	X	X	
(Tape Mark-Special Char.)	7-8	X			X	X	X	X
0	12-0	X	X	X	X		X	
A	12-1		X	X				X
B	12-2		X	X			X	
C	12-3	X	X	X			X	X
D	12-4		X	X		X		
E	12-5	X	X	X		X		X
F	12-6	X	X	X		X	X	
G	12-7		X	X		X	X	X
H	12-8		X	X	X			

FIGURE 211. 1401 CHARACTER CODE CHART IN COLLATING SEQUENCE

CARD LISTING (80-80)

```
10317920105ASD-TDR-63-144(PT.1)
103179202AMERICAN OIL CO. RESEARCH AND DEVELOPMENT
103179203DEPT, WHITING, IND
103179204$RADIATION CHEMISTRY OF TRIPLE BOND
103179205COMPOUNDS. PART 1. ACETONITRILE AND
103179206BUTYNE-2, LIQUID PHASES$
103179207T D NEVITT, R E RONDEAU, L A HARRAH ET AL
103179231FEBRUARY 1963  26P
103179251  (CONTRACT AF-33(616)-8247)
103179252R.1C.630304-MC
10217000105NRDL-TR-606
102170002NAVAL RADIOLOGICAL DEFENSE LAB, SAN FRANCISCO
102170003$EFFECTS OF DIVIDED DOSES OF X-RAY ON
102170004MORTALITY AND HEMATOLOGY OF SMALL AND LARGE
102170005DOMESTIC ANIMALS$
102170006G F LEONG, W G WISECUP AND J W GRISHAM
102170031JANUARY 7, 1963  22P
102170032R.2C.630225-MC
10316420105AD-286653
103164202QUARTERMASTER RESEARCH AND ENGINEERING COM-
103164203MAND, NATICK, MASS
103164204$ORGANOMETALLIC RESEARCH USING GROUP IV
103164205ELEMENTS$
103164206M C HENRY, J G NOLTES, W DAVIDSON ET AL
103164231JUNE 1962  8P
103164232  (PRESENTED AT THE 1962 ARMY SCIENCE CON-
103164233FERENCE HELD AT U S MILITARY ACADEMY, WEST
103164251POINT, N Y, JUNE 20-22, 1962)
103164252R.1C.630225-MC
```

serial no. — sequence # — tells the no. of catalog cards to print

```
123556701100SBORN, A. D.
123556702$1962. THE INDEXING OF SCIENTIFIC LITERATURE.
123556703SIXTY MILLION PAGES A YEAR. THE MASSIVE PROBLEM
123556704OF SCIENCE INDEXING AND ABSTRACTING.$
123556731AUSTRALIAN J. SCI., 24 (11)  P429-437
123556732  ABSTRACTED IN AMER. DOC., JAN. 1963  P 89
123657801105SALTON, GERARD.  1961.
123657802$INFORMATION STORAGE AND RETRIEVAL.$
123657803SCIENTIFIC REPORT NO. ISR-1 ON CONTRACT
123657804AF 19(604)8509, PROJ. 5632.  AFCRL-62-77.
123657805COMPUTATION LABORATORY, HARVARD UNIVERSITY,
123657806CAMBRIDGE, MASS.
123657807  152P
123657808  AVAILABLE AS AD-274816 FROM THE OFFICE OF
123657831  TECHNICAL SERVICES, WASH. D.C., FOR $2.75.
123657835  (ABSTRACTED IN AMER. DOC.--JAN., 1963, P89
```

"31" causes 2 blanks between the record type and the next digit when printing catalog cards.

Class Problem No. 9 THE DESIGN AND EVALUATION OF A FILE SYSTEM
Librarianship 245 (Intro. to Documentation)
Charles P. Bourne, Instructor

As an individual term project, each student will have the opportunity in this exercise to personally design and evaluate the performance of a file system. Each student will be given the same collection of material to form the data base for the file (90 abstracts from the April 1963 Literature Notes section of American Documentation). Each student will design his own classification or indexing scheme to do the subject analysis and description of this file material. Each student will actually implement his scheme with edge-notched cards (with codes and notation of his own design), and will perform the initial design (but not development or implementation) of tab card and computer versions of his file system.

The performance of each of the individual's systems will be tested and evaluated (by a class committee) on the basis of relevancy and recall. For this exercise, recall will be defined as the number of correct search responses divided by the number of search queries; relevancy will be defined as the number of irrelevant items recalled divided by the total number of items recalled. Both recall and relevancy will be expressed as percentages. This provides a relative measure of comparison of the performance of all the various systems developed by the class members. For testing purposes, each system will be given 100 search questions to answer (the same questions will be given to each system). Each of these search questions has been synthesized from the file items, and can reasonably be answered by one or more known file items. These questions will be posed to the file through the indexing scheme of each individual system, and scores (i.e. how many of the known items were recovered, and how much irrelevant material was delivered) will be recorded.

It is the intent of this exercise to provide the student with a means for using and testing the techniques that have been learned so far in the methods of subject analysis, notation, and evaluation procedures; as well as some of the characteristics of edge-notched cards, tab cards, and computer systems.

A formal written report will be required of each student to describe: (1) the classification or indexing system that was used; (2) the notation on the edge-notched card (include a template or sample card); (3) the results of the performance evaluation; (4) suggestions for changes that might have improved the performance of the system; (5) the initial system design for a tab card version of this system, including a description of the formats, layouts, and notation of the necessary cards; and (6) the initial system design for a computer version of this system.

The tests and evaluations will be made during the week of May 2. All of the card systems should be in operation by that time. The final written reports will be due the week of May 9.

Class Problem No. 9 FILE SYSTEM EVALUATION
Librarianship 245 (Intro. to Documentation)
Student _____

Question	Abstract Nos. Delivered by Search	Number of Successful Recalls	Number of Irrelevant Items	Total Number of Items Delivered by the Search
1. What are the information services of the National Academy of Sciences-National Research Council?				
2. What do you have on the use of an automatic recording telephone in a library?				
3. Is there a bibliography which would trace the communication channels in medicine?				
4. What are some commonly used methods for character recognition?				
5. What seven themophysical properties are searched for in Chemical Abstracts in a new method of literature search?				
6. What is the relationship between size of vocabulary possessed by people and the degree of their success in society?				
7. Where can one find information about the special library profession?				
8. What is there on techniques of automatic abstracting?				
9. What is National Cash Register Company's new microimage technique for storage and retrieval of document images?				
10. What techniques have been developed for automatic coding of natural English words?				
TOTAL		_____	_____	_____

Student _____

Question	Abstract Nos. Delivered by Search	Number of Successful Recalls	Number of Irrelevant Items	Total Number of Items Delivered by the Search
11. What kind of a system is the Aeronutronic File Search Evaluator?				
12. What method might be used for coding and recalling information in large capacity memory systems?				
13. What are current bibliographic practices in the field of medicine?				
14. What are some of the programs used in the ASTIA automated search and retrieval system?				
15. Is there a computer program that answers questions phrased in ordinary English about stored data?				
16. What problems have been encountered in developing machines capable of reading printed matter?				
17. Is there a program designed to abstract and evaluate data on electrical and electronic properties of materials from scientific literature?				
18. What are the library procedures for analyzing technical reports for information retrieval?				
19. For what are the ASTIA guidelines for cataloging and abstracting used?				
20. Are there any criteria for judging the adequacy of abstracts?				
TOTAL		_____	_____	_____

Student_____

Question	Abstract Nos. Delivered by Search	Number of Successful Recalls	Number of Irrelevant Items	Total Number of Items Delivered by the Search
21. A bibliography prepared for the IGY indexed how many articles?				
22. Is there an abstract on how a computer performs syntactic analysis?				
23. What do you have on a survey of current U.S. research and develcpment in scientific documentation?				
24. What is the Magnavue system?				
25. What information do you have on the MEDLARS project?				
26. The ASTIA descriptor term "documentation" is comparable to what AEC term?				
27. What do you have on bibliographic coupling in technical papers?				
28. What is Keyword-in-Context Indexing?				
29. What do you have on a system of parallel-search memories?				
30. What are some of the problems encountered in indexing a corpus of documents mechanically without the use of machines?				
TOTAL		_____	_____	_____

Student _____

Question	Abstract Nos. Delivered by Search	Number of Successful Recalls	Number of Irrelevant Items	Total Number of Items Delivered by the Search
31. What information is available on machine translation?				
32. What are the names of three journals published in Communist China which contain articles in the field of physics?				
33. What is being developed to keep researchers in the aquatic sciences up to date with their fields?				
34. Find a description of the IBM Walnut system.				
35. Can a computing machine read documents and, on the basis of selected clue words, decide in which subject category the documents belong?				
36. Where is there a guide to microforms in print?				
37. What are some of the solutions for solving the problem of duplication in the technical libraries of the Department of Defense?				
38. Any information on the Xerox 914 Copier?				
39. Is there a report available which summarizes the present problems of word searching?				
40. How do photo [illegible] coatings differ from photographic emulsions?				
TOTAL		_____	_____	_____

FINAL COMPOSITE SCORE:

RECALL= (Total No. of successful Recalls) / (Total No. of relevant items) = ( ) / ( ) = ____

RELEVANCY = 1 - [(Total No. of irrelevant items recalled) / (total No. items delivered)] = 1 - [( )/( )] = _____

Class Problem No. 9 FILE SYSTEM EVALUATION
Librarianship 245 (Intro. to Documentation)

Answer Sheet

Question	This Question is answered by	Question	This Question is answered by
1	39	24	53, 59
2	36	25	20
3	42,43	26	75
4	12	27	71, 73
5	13	28	16, 74
6	90	29	60
7	21	30	64, 76
8	67, 68	31	6, 7
9	58	32	5
10	69, 70	33	32
11	31	34	33
12	51, 59	35	83, 88
13	20	36	15
14	22	37	8
15	23	38	46
16	6, 12, 48, 49	39	48, 49
17	27	40	50
18	28		
19	79	Total number of relevant items = 54	
20	82		
21	84		
22	86		
23	14		

THIS SET CONTAINS ENTRIES IN BOLDFACE AS REQUESTED.

SAMPLE

Literature Notes

Contributions of abstracts from readers and suggestions of books and articles for review or inclusion in this bibliography will be welcome and are actively solicited. Volunteer abstracters and reviewers are needed, as well as people with special linguistic talents. All copies of reprints, reports, and correspondence for this bibliography should be addressed to Mr. Charles P. Bourne, Stanford Research Institute, Menlo Park, California. In order to increase its coverage, this section will include some abstracts copied or adapted from other publications. The American Documentation Institute is not able to supply copies of the publications abstracted or cited.

This particular issue inaugurates the practice of providing an author and corporate source index to the reviews, abstracts, and citations in the Literature Notes for each issue. The first portion of the serial number on each abstract (e.g. 4/63–27) gives the month and year of the issue of *American Documentation* in which the review, abstract, or citation appeared.

Author and Corporate Source Index

Documentation in General

4/63–2

Pioneering in Machine Literature Searching and Retrieval. 1961. Naval Ordnance Laboratory, White Oak, Md. NAVWEPS rept. no. 7388; AD–255 291. 22 pp. incl. tables. OTS price $2.60.

Presented at the Sixth meeting of the Council of Librarians, East Coast Navy Labs., held at the Naval Ordnance Laboratory, White Oak, Md., on 5 October 1960. These papers cover the programs of the American Society for Metals and the Naval Ordnance Laboratory in developing information retrieval programs. The first paper evaluates the experience at ASM in machine searching versus conventional library reference searching and finds machine searching to be more effective. The mathematicians explain the direct and indirect file for organizing library material for machine searches. Also discussed is the computer program adopted for the NOL Library. In the final paper, the Library's use of the computer and some of the advantages are set forth (author).

4/63–3

General Bibliography—Textbook for Library Institutes. 1960. Aerospace Technical Intelligence Center, Wright-Patterson Air Force Base, Ohio. AD–255 416. 34 pp. OTS price $3.60. (Trans. no. F–TS–9909/III from Obshchaya Bibliografiya — Uchebnik Diya Biliotechnykn Istitutov: 183–209, 1957.)

4/63–4

An Experimental Communication Center for Scientific and Technical Information. 1960. M. M. Kessler. Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass. Report on Contract AF 19 (604)5200. 20 pp. 4G–0002; AD–255 656. Repriced $0.50.

It is not clear that an elaborate system of information retrieval alone will add significantly to the flow of scientific information.

4/63–5

List of Communist Chinese Scientific and Technical Periodicals. 1961. Library Services Section, Air Information Division, Washington, D. C. Annotated bibliography. 63 pp., 247 refs. AID rept. 61-21; AD–255 784. OTS price $6.60.

An annotated bibliography of 247 titles is presented which reflects Communist China's periodical publishing activity in the subject fields of science and technology, with the exception of life sciences. The titles are arranged in alphabetical order according to the Wade-Giles romanization system and are followed by the Chinese characters and the translated titles. Some entries also carry the Communist P'in-yin or the Latin titles when these are available on the original publications. The annotation gives the frequency, place of publication, and publishing body of the journal, and a sketch of its characteristics with a brief analysis of its contents. Notes are added to indicate title changes and other pertinent data. The holdings of the Air Information Division are given. A subject index to the periodical titles is provided at the end of this bibliography (author).

4/63–6

Machine Translation. A Review and Analysis Report. 1961. Dimitri A. Kellogg. Office of the Chief, Research and Development, Army Research Office, Washington, D. C. 39 pp. ARO rept. no. 4; AD–257 839. OTS price $0.50.

A summary of the history, objectives, present status, and projected plans of the Army program in machine translation has been made. A similar review in less detail is included for all other government-sponsored research in machine translation. The Army program consists of three major efforts: machine translation of scientific literature from Russian to English, a similar program from German to English, and the development of a print reader capable of scanning Russian journal pages to furnish input for the machine translation. Examples of results of present efforts are included (author).

4/63–7

The Literature of Machine Translation: An Annotated Bibliography. Barbara Ann Bryce. Autonetics, Downey, Calif. Rept. no. EM–7249; AD–262 633. 272 pp. OTS price $17.50.

This annotated bibliography of 818 references presents most of the available literature on machine translation and shows progress made yearly from 1949 to mid-1961 in the MT field, both in the United States and abroad. The references are grouped alphabetically within these years. An appendix follows, listing approximately 40 individuals and organizations working on MT, with a brief description of their research and future plans in research and development of machine translation. Author and source indexes follow Appendix (author).

4/63–8

A Plan to Reduce Costs of Technical Library Operations in the Department of Defense. 1961. E. H. Langenbeck. U. S. Naval Ordnance Laboratory, White Oak, Md. Report NOLTR 61–102; AD–262 935. OTS price $1.60.

From a management point of view, unnecessary expenditures in duplication of cataloging technical reports exist in the technical libraries of the Department of Defense. This duplication can be eliminated by establishing DOD-wide standards for descriptive, subject, and subject code cataloging, and a standard code and dictionary for computer retrieval. The plan proposed presents the framework for solution of the problem and indicates the potential savings to be realized (author).

4/63–9

Report on Study of Status of Aerospace Corporation in Information Retrieval. 1961. Allen Kent, Jessica Melton, and Alan M. Rees. Center for Documentation and Communication Research, Western Reserve University, Cleveland, Ohio. AD–267 401. 254 pp. incl. illus. tables.

A study was conducted to review and analyze Aerospace's requirements relative to encoding, storing, and retrieving technical literature in an effort to determine the feasibility of applying automated (or other) techniques for improving control and service in the library — taking into consideration future requirements and conditions of maximum library

workload. A team from Western Reserve University made a diagnostic survey at Aerospace and other locations (ASTIA, Lincoln Laboratories, Battelle Memorial Institute, Mitre Corporation) and surveyed the literature resources and services in the United States of potential interest to Aerospace. Time was spent in reduction and analysis of survey results, and in the investigation of alternate systems. The Center for Documentation and Communication Research also maintained continuing contact with the various government agencies active in this field in order to insure awareness of new and proposed developments relating to the coordination of information services at the national level. The results of these investigations to date are a series of several stages of recommendations, the first for immediate implementation (author).

4/63–10

Theory of Documentation and Searching Strategy. 1962. Center for Documentation and Communication Research, Western Reserve University, Cleveland, Ohio. Final technical documentary report. 41 pp. Contract AF 49(638)357; AFOSR–2710; AD–278 551. OTS price $4.60.

4/63–11

A Survey of Computer Programs for Chemical Information Searching. 1961. Ethel C. Marden and Herbert R. Koller. National Bureau of Standards, Washington, D. C. 87 pp., 197 refs. NBS technical note 85; PB 161 586. Order from OTS $2.25.

The authors describe twelve computer systems for searching chemical literature. Preceding discussion of the computer systems, a brief description is given of different chemical notation systems, indexing and abstracting procedures, punched card systems (which were the forerunners of the computer systems), and special purpose literature-searching machines. A short discussion of the difficulties (linguistic and other) attendant to literature searching terminates the paper (author).

4/63–12

Automatic Character Recognition. A State-of-the-Art Report. 1961. Mary Elizabeth Stevens. National Bureau of Standards, Washington, D. C. 168 pp., 549 refs. Technical note no. 112; PB 161 613. OTS price $2.50.

A state-of-the-art report on current progress in automatic character recognition is presented. Areas of applicability and possibilities for controlled solutions to automatic character reading problems are discussed. Some commonly used methods for character recognition, the steps involved in a generalized recognition process, and comparative characteristics of certain representative character recognition systems are considered. Prospects for further progress, including potentially related research in pattern recognition, are reported (author).

4/63–13

Analytical and Experimental Study of a Method for Literature Search in Abstracting Journals. 1960. A. O. Cezairliyan, P. S. Lykoudis, and Y. S. Touloukian. Thermophysical Properties Research Center, Purdue University, Lafayette, Ind. 79 pp., 3 refs. TPRC rept. 11; PB 171 478. Order from OTS $2.00.

A new method for literature search using abstracting journals developed by Lykoudis, Liley, and Touloukian is extended. The method consists of searching an abstracting journal for a certain period of time and then obtaining new information by going though the bibliography section of the papers located by the direct search of the abstracting journal. This procedure can be repeated in cycles. According to the model of the new method of literature search, a direct search of *Chemical Abstracts* is made to locate publications on seven themophysical properties (thermal conductivity, specific heat, viscosity, diffusion coefficient, emissivity, thermal diffusivity, and Prandtl number) for all matter. This search is done for the period between January 1954 and December 1957. The references cited by each publication thus uncovered are searched to obtain further new information on the above-mentioned properties.

4/63–14

Current Research and Development in Scientific Documentation, No. 10. 1962. Office of Science Information Service, National Science Foundation, Washington, D. C. 383 pp. Catalog No. NS 2.10:10. $1.25.

The tenth in a series of descriptive reports on current research and development in scientific documentation, this publication, compiled by the Office of Science Information Service of the National Science Foundation, reports on all pertinent activities in the United States that have come to the attention of the Foundation staff, as well as foreign projects on which information could be obtained.

4/63–15

Subject Guide to Microforms in Print. 1962–1963. 1962. A. J. Diaz, Ed. Microcard Editions, Inc., Washington, D. C. 88 pp. $4.00

This is a comprehensive guide, by subject classifications, to materials (except theses and dissertations) which are available on microfilm and other microforms from U. S. publishers. It is the first comprehensive subject index to microreproduced materials. It is not a union list of microforms, but it is a classified listing of over 11,500 microform publications offered for sale on a regular basis. Each entry includes the price of the work, the publisher, and the method of microreproduction used. The guide covers the works of 47 publishers of information in microform. This guide will be published once a year and is a companion to the existing *Guide to Microforms in Print*, which lists microreproduced material alphabetically by author or title.

4/63–16

Information Retrieval. 1961. J. H. Veyette, Jr. *The American Behavioral Scientist*, Vol. 4, No. 10, pp. 15–10 (June).

The mass of information being published and the overlapping and interacting of scientific disciplines to create new technologies contribute to the current problem of information retrieval. The problem exists for the individual, the corporation, and the nation and has international significance as well. Information retrieval is defined, and some difficulties associated with it are discussed. Two recent developments of IBM research and experimentation are briefly described: Keyword-In-Context Index and Selective Dis-

semination of Information. Some characteristics of regional or discipline-oriented information centers linked together by a high-speed communication system are described (*IBM J.*).

4/63–17

A Survey of Languages and Systems for Information Retrieval. M. Grems. *Commun. ACM,* Vol. 5, No. 1, pp. 43–46 (January).

This survey is directed toward machine-oriented information retrieval languages and systems. It is the first published list for this area of interest to include the specific data items: application, equipment, source language, status, advantages, limitations, and contact person (*IBM J.*).

4/63–18

Problems Posed by an Expanding Technical Literature. 1962. Charles P. Bourne. *IRE Transactions on Engineering Writing and Speech,* Vol. EWS–5, No. 1, pp. 2–8 (August).

A quantitative estimate is made of the magnitude of the world's electronics and electrical engineering journal literature problem. By the use of a number of basic sources of information, a composite picture is established to show total volume, linguistic and national origins, degree of coverage by the abstracting and indexing services, and the availability of special information centers. Some criticism is directed at the electrical engineering professional societies for not taking more positive action on this problem of concern to their entire membership. The paper concludes with suggestions for coping with this general problem (author).

4/63–19

Concerning the Possibility of a Cooperative Information Exchange. 1962. M. Kochen and E. Wong. *IBM J.,* **6**(2): 270–271 (April).

In this note both the existing channels and potential channels of communication are considered. Theoretical arguments are advanced to establish the conditions under which a semi-automated information exchange system, advantageous to the participants, can be established. The principal factors to be considered and the parameters to be estimated prior to the implementation of such a system are discussed (author).

4/63–20

Stresses in Current Medical Bibliography. 1962. Frank B. Rogers. *New Engl. J. Med.,* 267:704–708 (October 4).

He notes the increase in medical journals; the present inadequacies of scientific publications; the growth in secondary publications, such as abstracting and indexing services, which seems more rapid than that of the basic literature; new techniques for indexing; the use of contents page listings; new methods of composition; changing bibliographic practices; and the MEDLARS project.

4/63–21

The Importance of Professionalism for the Special Librarian. 1961. M. Griffin. *Spec. Libraries,* Vol. 52, No. 8, pp. 462–463 (October).

The special librarian shares with other professionals in business and industry a dual loyalty — allegiance to his profession and to his company. He is obligated to uphold the standards learned during his training, to attain technical competence, and to grow on the job. The special library is company oriented; the nature and size of its collection may require that traditional ways of library organization be discarded or modified (*IBM J.*)

Total System Description, Design, or Evaluation

4/63–22

Evolution of the ASTIA Automated Search and Retrieval System. 1961. William E. Hammond. Armed Services Technical Information Agency, Arlington, Va. 71 pp., 3 refs. PB 915; AD–252 000. OTS price $2.00.

This report covers the evolution of the ASTIA automated search and retrieval system. Early consideration and punch card approaches are treated briefly. The current system and programs are described in detail with the exception of the creation of the Thesaurus which is covered in two other ASTIA publications — AD–227 000 and AD–247 000. The programs discussed are for the UNIVAC Solid State 90 magnetic tape system. Flow charts are included of the program now in use which executes ten simultaneous searches to four levels of coordination. Modification of this program to provide for either ten six-level coordinations or any combination of a 60-descriptor coordination is also discussed (author).

4/63–23

Baseball: An Automatic Question-Answerer. 1960. Bert F. Green, Jr., Alice K. Wolf, et al. Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass. 6 pp. AD–257 778. OTS price $1.10.

Baseball is a computer program that answers questions phrased in ordinary English about stored data. The program reads the question from punched cards. After the words and idoms are looked up in a dictionary, the phrase structure and other syntactic facts are determined for a content analysis, which lists attribute-value pairs, specifying the information given and the information requested. The requested information is then extracted from the data matching the specifications, and any necessary processing is done. Finally, the answer is printed. The program's present context is baseball games; it answers such questions as "Where did each team play on July 7?" (author).

4/63–24

The Evaluation of Information in Organizations. 1961. Roy Radner. University of California, Berkeley. 63 pp. Technical rept. no. 2; AD–262 550 (Contract Nonr–22277, Proj. NR–047–029). OTS price $6.60.

Contents:

1. Team decision problems with quadratic payoff functions.
2. Characterization of optimal decision functions and value of information in the quadratic case.
3. Generation of information structures by processes of observation, communication, and computation.

4. Complete communication, complete information, and routine.
5. No communication and a case of complete informational decentralization.
6. Partitioned communication.
7. Dissemination of independent information.
8. Error in instruction.
9. Complete communication of erroneous observations.
10. Management by exception: reporting exceptions.
11. Management by exception: emergency conference.
12. Comparisons among the several information structures.

4/63–25

Document Data Processing Central AN/GSQ–11A(V) (Minicard). 1961. Categories I and II test rept. N. Miullo. Rome Air Development Center, Griffiss Air Force Base, New York. 68 pp. incl. illus tables. Rept. no. RADC TR 61–137, in cooperation with Eastman Kodak Co., Rochester, New York, Contract AF 30(602)1698. AD–265 333. OTS price $7.60.

Results are presented of Categories I and II (AFR 80–14) testing of the Document Data Processing Central AN/GSQ–11A(V) (Minicard), under the cognizance of Directorate of Intelligence and Electronic Warfare, Rome Air Development Center (AFSC), Offutt Air Force Base, Nebraska, during the period 4 January through 10 March, 1961. The report describes test procedures; presents data on production, idle and down time; identifies and analyzes deficiencies in equipment and recommends solutions therefor; and, finally, discusses overall test results. As a consequence of this testing, it was determined that the AN/GSQ–11A(V) is operationally effective as a data-processing system (author).

4/63–26

Mathematical Models for Information Systems Design and a Calculus of Operations. 1961. Magnavox Research Laboratories, Torrance, California. 178 pp. incl. illus. tables. Final rept. MRL Rept. no. R–451; Contract AF 30(602)2111, Proj. 4594; RADC TR 61–196; AD–266 577. OTS price $13.00.

A general measure of component efficiency was investigated and applied to several types of information system components. Utilizing this measure of component efficiency, a model based on linear algebra was developed for a total information system. This model includes a generalized measure for total system efficiency. The application of the model was illustrated on a set of typical information storage, file organization, and equipment design problems. A calculus of operations, based on the measure of total system efficiency, was developed with broad application at all levels of system complexity (author).

4/63–27

Information Retrieval Program Electronic/Electrical Properties of Materials. 1961. Hughes Aircraft Co., Culver City, California. Quarterly progress rept. no. 1, July-Sept., 1961, by E. M. Wallace. 24 pp. incl. illus. Document no. 5171.2/8; Contract AF 33(616)8438; AD-270 500. OTS price $2.60.

A program was established to collect, index, abstract, compile, and evaluate data on the electrical and electronic properties of materials from scientific and technical literature. The program is divided into two phases. Phase I consists of determining and establishing the systems, procedures, and facilities for 1) searching, acquiring, abstracting, and indexing the literature; 2) storing, retrieving, compiling, and evaluating the data; and 3) publishing and disseminating bibliographies, property tables, data sheets, and summary reviews of important topics. Phase II consists of performing the above tasks for two categories of materials: semiconductors and insulators (author).

4/63–28

Procedures for Analyzing Technical Reports for Machine Retrieval. 1962. Eva Liberman. Naval Ordnance Laboratory, White Oak, Md. 29 pp. incl. illus. NOLTR 62–51; AD–278 454. OTS price $3.60.

Library procedures for analyzing technical reports for information retrieval are discussed. Identification information, subject analysis, IBM card-punching and search procedures have been established to meet the needs of the Laboratory for its collection of technical reports (author).

4/63–29

VIP. A Variable Information Processing System for Storage and Retrieval of Missile Data. 1962. M. Kosakoff and D. D. Buswell. Naval Ordnance Laboratory, Corona, California. 29 pp. incl. illus. Technical memo. no. 64–474; AD–278 835. OTS price $3.60.

The Variable Information Processing (VIP) system for the storage and retrieval of missile data is described. The storage of information in VIP format permits the flexible application of combinations of general-purpose processors for the accomplishment of retrievals and summarizations (author).

4/63–30

General Information Manual: Selective Dissemination of Information. 1962. Brochure E20–8092 of the IBM Corporation, Data Processing Division, White Plains, New York.

4/63–31

Aeronutronic File Search Evaluator: An Electronic System for Automated High-Speed Information Retrieval. 1962. Technical brochure No. U–1587 from Aeronutronic, Newport Beach, Calif. 24 pages.

The Search Evaluator was developed under an Air Force contract and has the military designation of AN/GSQ–38. It is a digital system for searching and retrieving information from magnetic tape files. A special-purpose digital computer is employed to perform high-speed comparisons and evaluations at full magnetic tape speeds. Additional computer operations aid in monitoring the file records. The equipment consists of:

1. A magnetic tape transport and control unit.
2. A comparator and logic evaluator for processing the information from the magnetic tape.
3. A magnetic core memory for storing each record read from the magnetic tape.
4. An electric typewriter with a paper tape punch and reader to generate new inquiries or search criteria and to print out records which satisfy the search criteria.
5. A Biax memory for storing the search criteria.

The tape transport runs at the speed of 150 inches per second during the search, except during printout.

4/63-32

A Developing Aquatic Sciences Information Retrieval System. 1962. Joel S. O'Connor and Saul B. Saila. *Trans. Am. Fisheries Soc.*, **91**(2):151–154 (April).

The Aquatic Sciences Information Retrieval Center is being developed to meet the urgent necessity for rapid synthesis and dissemination of information related to the aquatic sciences. The inability of researchers to maintain familiarity with recent work in their discipline is looming as one of the most significant problems in many fields of endeavor, including fisheries and related aquatic sciences. Further rapid progress in the aquatic sciences is believed to be contingent upon an objective means of relieving this problem (author).

4/63-33

Photo-Image Storage — Its Role in Modern Business. 1961. John H. Veyette, Jr. *Business Automation*, Vol. 6, No. 4, pp. 16–21 (October).

This article describes the IBM WALNUT photo-image and retrieval system which provides storage capacity for millions of printed or typed pages and drawings, any one of which can be retrieved within five seconds. A hypothetical corporation (Better and Best Electronics Manufacturing Company) is used to show how the system would benefit R & D, Materials, Reliability and Environmental Test, Manuals and Handbook Preparation, and Manufacturing Control functions. It is pointed out that WALNUT was designed for a specific purpose and is not commercially available. However, the technology employed in the system would be applicable for informational support in many corporations with problems similar to the ones presented by BBEM Co. Photo-image technology shows promise of becoming a powerful tool for the storage and retrieval of information (*IBM J.*).

4/63-34

Transmitting Technical Information by Strip Microfilm. 1962. C. W. Carlson. *Graphic Sci.*, Vol. 4, No. 3, pp. 13–14 (March).

A system for reproducing and distributing engineering change documents is described, which utilizes a 35 mm planetary camera, a roll film xerographic printer, and film strips. The basis of the system is a procedure whereby engineering change documents are logged in, exposed on microfilm, and the microfilm is processed on an automatic processor. The engineering change is then incorporated on the master drawing, which is microfilmed and filed on an aperture card for later use. The system of logging in and processing the engineering changes is reviewed, and the role of the factory Blueprint and Reproduction Services groups is discussed. This entire procedure is described in detail, and the distribution of the documents throughout IBM is indicated. Flow charts of the process are included, along with a description of quality control measures utilizing gray scale charts, as well as lettering and spacing criteria for the original drawings (*IBM J.*).

4/63-35

Data Processing Techniques for Parts Cataloging. 1960. P. F. Santarelli and R. F. Cunningham. *STWE Rev.*, Vol. 7, pp. 4–7 (April). (Adapted from a paper presented by the authors at the 1958 Technical Writers' Institute, Rensselaer Polytechnic Institute.)

The computer techniques described in this paper are for the preparation and maintenance of an illustrated parts breakdown manual. These procedures are the culmination of a project undertaken at IBM in 1955 which resulted in the preparation of an illustrated parts breakdown manual consisting of more than six thousand pages of listing and fourteen hundred illustrations (*IBM J.*).

4/63-36

Disseminating Current Information. 1962. Sandra Hocken. *Spec. Libraries*, Vol. 53, No. 2, pp. 93–95 (February).

An effective method of sifting useful information from the mass of incoming material has been devised by the IBM Advanced Systems Development and Research Library in San Jose. A daily library newspaper lists new books and current articles of interest to the laboratory, chosen by volunteer scanners. Copies of microfilmed articles may be requested by calling an automatic recording telephone in the library (*IBM J.*).

Information Services

4/63-37

Technical Information and Bibliographical Centers in the USSR (Tekhniko-Informatsionnye i Bibliograficheskiye Tsentry v SSSR i Izdavayemye Imi Materialy). 1961. Ye. N. Morozova. Aerospace Technical Intelligence Center, Wright-Patterson Air Force Base, Ohio. 31 pp. Trans. no. MCL-796 of Obshchestvo Po Rasprostraneniyu Polsticheskikh i Nauchnykh Znaniy RSFSR; Leningradskiy Dom Nauchno-Tekhnicheskoy Propagandy, Leningrad: pp. 1–23, 1959. AD-259 239. OTS price $3.60.

4/63-38

Information Retrieval at the Missile Safety Information Center. 1961. R. C. Turner, Jr. Naval Weapons Laboratory, Dahlgren, Va. 49 pp. incl. illus. NWL rept. no. 1762; AD-259 508. Supersedes Technical memo. no. W-4/60 and Rept. no. 1724, AD-245 809. OTS price $5.60.

Information retrieval is an important function of the Missile Safety Information Center at the Naval Weapons Laboratory, Dahlgren, Virginia. The use of this retrieval system permits rapid and low-cost searching of the publications contained in (or referenced by) the Missile Safety Information Center. The search for discrete information is limited to those characteristics under which each document has been cataloged. The system is comprehensive and flexible. This report includes the categories of information used in the Information Retrieval system, details of the operation, and a description of the codes utilized. This report cancels and supersedes NAVWPNLAB, Dahlgren Technical Memorandum No. W-4/60, dated February 1960, and Report No. 1724 (AD-245 808), dated 27 October 1960. The superseded technical memorandum and report should be destroyed (author).

4/63–39

Scientific Information Activities of the National Academy of Sciences-National Research Council. 1962. Publication 1031. NAC-NRC, Washington, D. C.

Most of the information activities, except those relating to original research publications, are described for each of the eight divisions of the Academy-Research Council (Anthropology and Psychology; Biology and Agriculture; Chemistry and Chemical Technology; Earth Sciences; Engineering and Industrial Research; Mathematics; Medical Sciences; Physical Sciences). Several nondivisional activities (Geophysics Research Board; NAS-NRC Library; Office of Documentation; Pacific Science Board; Space Science Board; etc.) are also included. The date of origin (and termination), purpose and indication of financial support, the literature collection, information services, and publications (both published and planned) are given for each project or committee whenever appropriate. There is no overall subject or publications index, but the arrangement and size of the publication (50 pages) permit easy scanning.

User Requirements

4/63–40

Suggested Procedure for Study and Analysis of Information Requirements. 1961. Raymond Sabeh. Air Force Electronic Systems Division, Bedford, Mass. 8 pp. incl. illus. Rept. no. ESD TN 61–58; AD–260 352. OTS price $1.10.

The analysis stresses that critical to the design of Command-Control Systems is the determination of information requirements and knowledge concerning the manner in which the information is to be used operationally. A suggested procedure for study and analysis of information requirements is discussed. The procedures emphasize the identification of the systems missions, formulation of operational requirements and conditions under which the systems will perform, and analysis of the functions involving the organization of both man and machine directed at achieving the objectives prior to considering the available hardware or men who will serve as components (author).

4/63–41

The Scientist Who Knows Everything (Uchenyi Kotoryi Znaet Vse). 1962. R. Bershadskii, trans. by Z. Jakubski. Space Technology Laboratories, Inc., Redondo Beach, Calif. 41 pp. Rept. no. 9990–6138–KU–000; Trans. from *New World* (Novyi Mir), Moscow, Vol. 38: 177–197, April 1962. AD–278 503. OTS price $4.60.

4/63–42

An Annotated Bibliography of Studies on the Flow of Medical Information to Practitioners. Part I. 1961. Walter E. Boek. A report of the Institute for Advancement of Medical Communication, New York, New York. 27 pp., 80 refs.

This annotated bibliography reviews the literature that describes the results of studies of how the medical practitioner utilizes a great variety of available communication channels to learn about new developments, and what factors determine his preferential uses of some channels over others. This report would be very useful to those interested in the problems of assessing the effectiveness of efforts to serve the users' information needs. Reports of studies of methodology or of user groups other than physicians are not included.

4/63–43

An Annotated Bibliography of Studies on the Flow of Medical Information to Practitioners. Part 2. 1962. Walter E. Boek. A report of the Institute for Advancement of Medical Communication, Bethesda, Md. 8 pp., 17 refs.

This is a supplement to the original bibliography published by Boek in 1961. The type of material covered and the criterion for selection of the references for this supplement are the same as for the original report.

Document-Handling Methods

4/63–44

Cooperative Storage Warehouses. 1962. Helen J. Harrar. 203 pp. incl. tables, refs. Doctoral thesis. Contract AF 49(638)849; AFOSR–2615; AD–278 712. OTS price $14.00.

Specific Hardware

4/63–45

Image Format Considerations in Photographic Intelligence Processing. Metricard Design. 1961. A. W. Richmond. Thompson Ramo Wooldridge, Inc., Canoga Park, Calif. 55 pp. incl. illus. table. DSPO technical note no. 14; Contract AF 30(602)1814; RADC TN 61–63; AD–258 149. OTS price $6.60.

The significance of the format of photos, maps, charts, and other graphic data is discussed in terms of the effects on efficiency of photo data reduction systems. The varied functions in such systems suggest many different format-defining parameters. These parameters and their interrelationships are considered for the purpose of identifying formats of maximum utility for all such functions. It is concluded that a wide range of photo data reduction activities can effectively be accomplished with one standard size of unit record prepared from 70 mm roll film and containing an image with a minimum dimension very close to 2.25 inches. The other dimension of the image may require variation from 2.25 inches to a value about one-third larger to meet various geometrical and technical considerations. A description of such a standard format (Metricard) is presented. This format was designed to meet the requirements of photo data reduction systems emphasizing the features of rapid handling, moderate storage volume, and high information density (author).

4/63–46

The Development of a Minicard Viewer Processor MX–1993 (XW–1)/G. 1961. Xerox Corp., Rochester, N. Y.

Final engineering rept. 70 pp. incl. illus. tables. Contract AF 30(602)1523; RADC TR 61-128; AD-260 073. OTS price $6.60.

A proprietary device, to be known as the 914 Office Copier, was developed. A modified design of this unit provided the printout portion of the contract machine. Minicard-handling techniques developed by the Eastman Kodak Company were adapted and tailored to the needs of the developmental machine. An optical system was conceived and constructed, using bench-test technique with mock-ups. Various light sources were tested in the viewing-projection system, and a compromise selection was made on the basis of stability, light intensity, and generated heat. A system of scan exposure of the Minicard was developed, involving synchronized scanning of a Minicard with enlargement exposure on a revolving xerographic drum. The optical system was arranged to provide visual inspection by the operator of the 12 documents on the Minicard, enlarged 60 times on a translucent viewing screen. Rapid selection of documents was provided by a single 12-position switch and printout of any selected document by a push-button switch. Only one control was made manual, and it was equipped with positive positioning detents to avoid a requirement for a high degree of operator skill or dexterity (author).

4/63-47

Development Test Support for Minicard Viewer Processor MX-1993 (XW-1)G. 1961. Xerox Corp., Rochester, N. Y. Final technical rept. 51 pp. incl. illus. Contract AF 30(602)2343, Proj. 4591; RADC TR 61-197; AD-263 593. OTS price $5.60.

4/63-48

Word Searching Device. Part II. 1961. Thompson Ramo Wooldridge, Inc., Canoga Park, Calif. 49 pp. incl. illus. tables. Technical rept. no. C117-1U14, pt. 2; Contract AF 30(602)2300; AD-264 440. OTS price $4.60.

Additional work directed towards recognition of complete words is reported. It is shown that such systems are unattractive because of the extremely severe tolerance requirements to effect a satisfactory match. The basic conclusion of the report is that a word-searching device utilizing letter-by-letter recognition appears most promising. The system would have four basis units:

1. A text scanner incorporating a high resolution flying spot scanner and a suitable electronic servo to permit following lines of text.
2. A letter recognizer with associated logic having a learning capability.
3. A word detector utilizing a transparency that can be marked by grease pencil to the code of the desired key words and associated scanner, photo-multiplier and logic to detect occurrences of the key words.
4. A read-out recorder to display the occurrences of the key words for study to determine possible interest.

A simplified version of the word searcher is described which would show feasibility of all of the novel elements of the proposed system (author).

4/63-49

Word Searching Device. Part I. 1961. Thompson Ramo Wooldridge, Inc., Canoga Park, Calif. 84 pp. incl. illus. tables. Technical rept. no. C117-1U4, pt. 1; Contract AF 30(602)2300; AD-264 441. OTS price $8.10.

Considerable effort has been directed towards determining the basic limitations of a word-searching device due to the characteristics of printed text. This analysis has been aided by simulation studies based on correlation measurements made on enlargements of typical English characters and subsequent experiments in which a digital computer simulated the Word Searcher. This investigation has shown that the differences in output signal between the desired key words and possible false alarms may be quite small, with the result that the equipment tolerances must be closer than originally anticipated if successful operation is to be achieved. This report presents a detailed account of the investigation to date and of the difficulties encountered. The unexpectedly stringent tolerance requirements have led to modification of the original approach. Alternate mechanizations are considered which bypass the fundamental limitations of the flying-spot scanner system (author).

4/63-50

Theoretical and Experimental Investigation of Photochromic Memory Techniques and Devices. 1961. National Cash Register Co., Hawthorne, Calif. Final rept. 503 pp. incl. illus. tables, refs. Contract AF 33(616)6205, Projs. 7062 and 70921; ASD TR 61-70; AD-273 512.

Results are presented on the basic sensitivity and resolution capabilities of photochromic materials. Photochromic coatings are similar to photographic emulsions in that they exhibit a change in transmission with excitation by light of the proper spectral distribution. Photochromic coatings differ, however, in two respects: they do not require a development process and they exhibit a reversal of state (return to the first state) when illuminated by light of a second spectral distribution. In this reversal of state, photochromic coatings are similar, superficially at least, to magnetic coatings. Temperature and fatigue characteristics are also discussed, with special attention given the stability of the colored state with temperature. Chemical data is presented on the light-sensitive materials, the solvents, and the coatings.

4/63-51

Organization of Large Memory Systems. 1962. R. S. Ledley. In *Large-Capacity Memory Techniques for Computing Systems*, M. C. Yovits, Ed., The Macmillan Company, New York, pp. 15-51.

Methods for recall of information from memory systems can be based on the information's location or address in the memory, which currently is most common in electronic digital systems, or on the content or substance of the information, as currently used in indexing systems. Either technique can be built into the computer as the basis for the memory systems organization.

Location-based recall methods are denoted as *addressing* and can be classed as *simple* and *complex*. Simple addressing covers the usual *one-at-a-time* word selection, *two-or-more-at-a-time* word selection, and *multiple-operational-unit* memory-word selection. Complex addressing includes *partial addressing*, which results in possible variation of word

length; *indirect addressing,* to enable look-up-table memories and larger memories than could be directly addressed; and *chain addressing,* for facilitating push-down-list retrieval frequently required for automatic programming techniques.

Content-based methods for recall of stored information are called *contenting* (for want of a better term) and can be classed as *simple* and *complex.* Simple contenting includes the *exhaustive search,* where the entire memory is scanned effectively; the *linked search,* where successive searches are made, each depending on the previous outcome; and the *directed search,* where successive localization is accomplished in an iterative manner. Complex contenting involves *coordinate contenting,* where the information is organized in repetitious form to facilitate multiple coordinated access; *conditional contenting,* where alternative logical conditions are placed on the contenting process; and *combined contenting,* which includes addressing as well as contenting. Finally, the promise of a billion gates leads to systems where the logical design and the memory functions merge indistinguishably (author).

4/63–52

Large Files for Information Retrieval Based on Simultaneous Interrogation of All Items. 1962. J. Goldberg and M. W. Green. In *Large-Capacity Memory Techniques for Computing Systems,* M. C. Yovits, Ed., The Macmillan Company, New York, pp. 63–77.

There is increasing interest in memories in which items are retrieved on the basis of content rather than location and in which the tests of content proceed simultaneously for all items. A study is being made of the usefulness of this method for realizing a large document-retrieval index file. This application permits use of semipermanent storage but requires that the cost per bit of storage be very low and that the equipment be capable of separately identifying each of an unknown number of responses to a given search criterion. An efficient testing method has been found for resolving multiple responses which permits a construction employing only one sensing element for an entire memory.

Feasible tests for searching are "identity (partial or complete)," relative size, i.e., "greater-than" or "less-than," and "inclusion." The latter is appropriate to coding schemes in which description terms are superimposed in a common field. At the price of certain tolerable inconveniences, this coding makes possible significant simplification in field searching and economies in bit storage. The early studies of this coding by Mooers and Wise have been refined, and the properties of a certain model have been calculated numerically.

Realizations in several technologies have been investigated, and several examples will be presented employing magnetic, electro-optical, and cryogenic elements, respectively. The geometries of these realizations have been designed to facilitate insertion of the semipermanent data (author).

4/63–53

Combined Magnetic and Graphic Store. 1962. R. L. Laurent. In *Large-Capacity Memory Techniques for Computing Systems,* M. C. Yovits, Ed., The Macmillan Company, New York, pp. 137–147.

A new concept in information storage and retrieval, Magnavue, combines graphic and magnetic information on a single 1-inch by 3-inch card. Each card has provision for a photographic image and 450 magnetically recorded alphanumeric characters identifying and describing the image. The image may be of a document, chart, graph, fingerprint card, aerial photograph, or any other item that contains information to be retrieved.

The complete Magnavue system includes:

1. The card transport and display unit, which reads and transports the cards at rates up to 90 a second. This unit can project the image on the local display station, send the card to a remote display station, and/or produce a hard-copy output.

2. Rapid-access files providing an on-line storage for up to 900,000 cards.

3. The electronic control unit, which controls the operation of the files and the card transport and display unit.

4. A central processor, which is the overall intelligence for the system. It stores information to be written on the cards and commands for the electronic control unit.

The cards are prepared in five steps. First the document is photographed, and then the image is processed and transferred to the card. The card is then laminated and cut to the proper format. Last, the magnetic information is written on the card by the Magnascriber, a desktop keyboard device.

The wide field of application for Magnavue includes document storage, photo reconnaissance, briefing room display, and remote display (author).

4/63–54

The N. C. R. Magnetic Card Random-Access Memory. 1962. A. M. Angel. In *Large-Capacity Memory Techniques for Computing Systems,* M. C. Yovits, Ed., The Macmillan Company, New York, pp. 149–162.

A good definition of an ideal random access memory might be "a memory of infinitely large storage instantaneously accessible and economically feasible," but this definition would describe nothing even near existent in the data-processing field. Actually, those random access memories presently in use, or contemplated for the relatively near future, are, in reality, quasi-random access machines. They have been derived in general from two devices well-known in the computer art — the magnetic drum and the magnetic tape file — through almost ten years of evolution.

One of the most recent and unusual of these devices is the CRAM (NCR Magnetic Card Random Access Memory). Contained in CRAM are 256 magnetic oxide-coated cards, any one of which may be accessed in less than one-fifth of a second and reaccessed in 46 milliseconds. These cards are 3¼ inches wide and 14 inches long, and the 256 of them contain over 5½ million alphanumeric characters. Further, the card may be removed and replaced with greater simplicity than a reel of magnetic tape.

The most important significance of CRAM is that it is the first storage device capable of covering the full spectrum of the application from magnetic tape file through random access.

The salient features of these new concepts are considered. A complete description of the mechanism follows, while the

control logic and features of the electronics are touched upon lightly (author).

4/63–55

Data Processing with the Photostore. 1962. G. W. King. In *Large-Capacity Memory Techniques for Computing Systems,* M. C. Yovits, Ed., The Macmillan Company, New York, pp. 301–304.

It is the purpose of this paper to report on the development of the Photostore, as the original feasibility studies were supported by a contract with the Office of Naval Research. The prototype machine which resulted encouraged the Department of the Air Force (RADC) to support the construction of a full-scale machine with input and output to provide a complete operating system (author).

4/63–56

A Large-Capacity Document Storage and Retrieval System. 1902. R. W. Porter. In *Large-Cpacity Memory Techniques for Computing Systems,* M. C. Yovits, Ed., The Macmillan Company, New York, pp. 351–360.

A large-capacity, random access document retrieval system providing greatly compressed image storage of many millions of documents and rapid selective output is described. This system, now in advanced development, consists basically of a random access document index, a document input converter, and a random access image file.

The system provides for the storage of 990,000 pages per image-file module, and these modules can extend the system capacity indefinitely. The high-speed, high-capacity random access magnetic index provides great flexibility in recording or searching records of documents in 1 to 2 seconds; the large-capacity, random access document image files provide reproduction of any desired document in 5 to 10 seconds (*IBM J.*).

4/63–57

Rotating-Mirror Photographic Storage Systems. 1962. D. M. Baumann. In *Large-Capacity Memory Techniques for Computing Systems,* M. C. Yovits, Ed., The Macmillan Company, New York, pp. 373–383.

A rotating mirror photographic storage system, here called "Photomemory," is a high-speed device for retrieval of binary information. The memory output is in the form of parallel binary channels. The binary word length is dependent on the number of photoreceptors used. In the present systems the maximum reading rate is dependent only on the response time of presently available detectors and their associated circuitry. The photomemory is essentially a sequential access device, but its potential high scanning rate allows it to be considered for pseudorandom search; i.e., if the scanning is rapid enough, the time required to read a word or group of words may compare favorably with the time required in random access devices.

The storage media of these devices are photographic slides of ordinary "press camera" proportions. The first slides of each kind are prepared on the photomemory device, but succeeding copies can be reproduced by ordinary photographic techniques. Each slide can contain the approximate equivalent of a reel of magnetic tape. A random access slide changer can give the system access to an entire library of information.

The test results of the feasibility demonstration models, built at M.I.T., are discussed, along with a brief presentation of the results of the optimization of the geometrical design parameters of the photomemory system. Present and projected applications of the photomemory systems and the utilization of rapid developing and erasable memory patterns are also discussed (author).

4/63–58

The Photochromic Microimage Memory. 1962. C. O. Carlson, D. A. Grafton, and A. S. Tauber. In *Large-Capacity Memory Techniques for Computing Systems,* M. C. Yovits, Ed., The Macmillan Company, New York, pp. 385–410.

Recent research and development within The National Cash Register Company has resulted in a new technique for the storage and retrieval of document images. This technique has made very high density document storage practical at reductions of 200:1 and greater. Miniaturization of document image size to this degree makes feasible a relatively simple, high-speed, random access capability. Writing and storage rates of 7,200 image fields per hour serially and 4,300 per hour randomly are envisioned. In the retrieval mode of operation, when recording the output on film, rates of 36,000 per hour serially and 10,000 per hour randomly are expected. In addition to the ability to create static photomemories, this technology provides a new capability in document manipulation by its ability to alter the stored document image by erasure and rewriting. The equipment described in this paper provides one form of photochromic mechanization and, with respect to hardware, a number of alternate approaches are possible. However, a specific configuration leads to a clearer discussion of details than would be possible with the more general approach. The system to be discussed has not been built, beyond a breadboarding of some of the functions. The laboratory studies have indicated that this system is not only feasible but well within the state-of-the-art. Examples of output copy from this type of memory system are shown in the Appendix (authors).

4/63–59

Large-Capacity Memory Techniques for Computing Systems. 1962. Edited by M. C. Yovits. The Macmillan Company, New York. (Based on the symposium on Large-capacity memory techniques for computing systems sponsored by the Information Systems Branch.)

4/63–60

Algorithms for Parallel-Search Memories. 1962. A. D. Falkoff. *J. ACM,* Vol. 9, No. 4, pp. 488–511 (October).

The underlying logical structure of parallel-search memories is described; the characteristic operation of three major types is displayed in the execution of searches based on equality; and algorithms are presented for searches based on other specifications, including maximum, minimum, greater than, less than, nearest to, between limits, and ordering (sorting). It is shown that there is a hierarchy of dependency among these algorithms; that they appear in pairs, with each member of a pair belonging to one or the other of two distinct classes; and that every type of search can be executed within each class (author).

4/63–61

Table Look-Up Procedures in Language Processing. Part I. The Raw Text. 1961. G. W. King. *IBM J.*, Vol. 5, No. 2, pp. 86–92 (April).

A method of addressing memories is described which is very powerful in the processing of natural languages, where the arithmetic or logical operations are either nonexistent or do not lend themselves to algorithmic description. The main feature is the guarantee of initiation of an exhaustive search for a linguistic word at a point just beyond the desired address. Sequential search backwards not only locates an address if it is there but also provides identification of a longest match first. The method is further extended to provide "conditional" addressing by prefixing subsequent addresses from information obtained in earlier searches (author).

4/63–62

Associative Memory with Ordered Retrieval. 1962. R. R. Seeber and A. B. Lindquist. *IBM J.*, Vol. 6, No. 1, pp. 126–136 (January).

A basic associative memory utilizing cryogenic circuitry is described, and its functions are compared with those of previously published associative memory descriptions. The ordered-retrieval sorting algorithm is described, along with its implementation by means of a ternary interrogating counter. A sorting example is given. The sorting efficiency is discussed, and an efficiency formula is given. The required additions to the basic memory are outlined. Finally, some of the basic cryotron circuits are illustrated, and their operation is described (author).

4/63–63

A Magnetic Associative Memory. 1961. J. R. Kiseda, H. E. Petersen, W. C. Seelbach, and M. Teig. *IBM J.*, Vol. 5, No. 2, pp. 106–122 (April).

This paper describes a computer storage system in which data flows in and out of the memory on the basis of content rather than location (address). In addition, a small experimental model of this system is described, using ferrite cores as novel associative memory storage elements (author).

Techniques and Reference Tools for Subject Analysis and Description

4/63–64

Some Suggested Mechanized Indexing Investigations Which Require No Machines. 1960. John O'Connor. Institute for Cooperative Research, University of Pennsylvania. 19 pp., 13 refs. Rept. on Contract Nonr-551(35); PB 149 077; AD-240 040. OTS price $1.60.

An investigation was made concerning the possibilities of mechanized indexing which would not require machines. The equipment required for the investigations is a vocabulary of indexing terms and a document collection to which they have been applied and are being used for retrieval. Some rules were postulated for the selection and assignment of terms. The underassigning and overassigning of terms was discussed at length.

4/63–65

On Formal Properties of Simple Phrase Structure Grammars. 1960. Y. Bar-Hillel, M. Perles, and E. Shamir. Hebrew University, Israel. 61 pp., 8 refs. Technical rept. no. 4 on Contract N62558-2214; PB 150 192; AD-241 572. OTS price $6.60.

Mathematical problems are presented concerning simple phrase structure grammars (SPGs). The problems studied include representations of SPGs, relations with other formal systems (in particular with finite automata), characterizations of the language obtainable by SPGs, closure under Boolean operations, and decision problems for various properties of SPGs (author).

4/63–66

Research in Inductive Inference. 1961. R. J. Solomonoff. Zator Co., Cambridge, Mass. 21 pp., 4 refs. Progress rept. for 1 April 1959–30 November 1960 on Contract AF 49(638)376. ZTB–139; AFOSR–160; AD–252 446; PB 155 215. OTS price $2.60.

The principal progress made was in the discovery of what are apparently several equivalent formal solutions to the general inductive inference problem. These solutions are applicable to numerical and/or non-numerical and/or analog and/or digital data. Any type of information that is available can be made part of the evidence upon which the inferences are made. These general induction methods were applied to several specific problems in non-numerical prediction. Some computer programs were written for the discovery of regularities in English text and any other sequence of symbols. Some work was done toward programming a computer to learn to assign descriptors to documents. Before the new inference methods were discovered, much time was spent on the problem of discovering the grammars of phrase structure languages from a body of text alone. Since then, the general inference methods have cleared up a serious point of difficulty in this problem. Another problem upon which considerable progress was made is the problem of programming a computer to improve its own inference methods (author).

4/63–67

The Study for Automatic Abstracting C107–1U12 (Appendix D). 1961. Thompson Ramo Wooldridge, Inc., Canoga Park, Calif. Final rept. 1v incl. illus. tables. Contract AF 30(602)2223; RADC TR 61–230, vol. 2; AD–269 599.

4/63–68

The Study for Automatic Abstracting C107–1U12. 1961. Thompson Ramo Wooldridge, Inc., Canoga Park, Calif. Final rept. lv. incl. illus. tables. Contract AF 30(602)2223; RADC TR 61–230, vol. 1; AD–269 600.

Techniques for the automatic abstracting of textual information are presented. One approach is the study of attributes of controlled human-produced abstracts to deduce from such study a mechanizable technique for creating similar products. A procedure making efficient use of mechanized data handling was implemented to collect, organize, and present language data for analysis. The results of such analysis were fed back to the machine as the basis for a

new cycle of experiments. A second approach is based on a sentence classification scheme for content representation as a measure for determining the relative significance of different portions of an article. Extract-type abstracts were produced manually for 100 articles; each article was extracted separately by three individuals. Each sentence was coded to reflect the number of persons selecting it for extraction. Frequency counts of the material in each category were made. The full text of the sample library, together with the above information, was recorded on punched cards, and a computer program was written and run. The second major approach consists of theoretical investigations to code, identify, and categorize the recognizable and appropriate attributes of sentences, paragraphs, and articles (author).

4/63–69

Automatic Word Coding Techniques for Computer Language Processing. 1962. William R. Nugent and Alexander Vegh. Itek Corp., Waltham, Mass. Final rept. lv. incl. illus. tables, refs. Rept. no. IL 9018–1, vol. 1; Contract AF 30(602)2377; RADC TDR 62–13, vol. 1; AD–272 401. OTS price $13.50.

This analysis concerns the automatic coding of natural English words. The objective is the derivation of coding techniques particularly suitable for representation, manipulation, and storage in computers. Several classes of techniques were explored and tested with large samples of words. There have evolved some contributions to a general theory of word coding and three broad classes of coding techniques (author).

4/63–70

Automatic Word Coding Techniques for Computer Language Processing. Sample Results of Computer Tests. 1962. William R. Nugent and Alexander Vegh. Itek Corp., Waltham, Mass. Final rept. lv. Rept. no. IL 9018-1, vol. 2; Contract AF 30(602)2377; RADC TDR 62–13, vol. 2; AD–272 402. OTS price $13.50.

4/63–71

An Experimental Study of Bibliographic Coupling Between Technical Papers. 1961. M. M. Kessler. Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass. 10 pp. incl. tables, 3 refs. (In cooperation with National Science Foundation.) AD–278 226. OTS price $1.60.

A single item of reference shared by two papers is defined as a unit of coupling between them. Based on this unit, two graded criteria of coupling are defined. These are applied to a test population of papers to generate smaller related groups. The process was tested manually with results that encourage more extensive experimentation (author).

4/63–72

The Effects of Formatting Restrictions on the Quality of Report Summaries Produced by Humans. 1962. John D. Ford, Jr. System Development Corp., Santa Monica, Calif. 10 pp. incl. illus. tables. Rept. no. SP–707; AD–278 663. OTS price $1.60. (Presented at the American Psychological Association Meeting, St. Louis, Missouri, September 5, 1962.)

Despite increasing automation of information-retrieval systems, humans will still be required to assist in selecting and reducing information from text. Format restrictions imposed by man-machine communication might interfere with verbal habits and result in performance degradation. However, such restrictions might result in increased effectiveness, because of partial encoding. A restricted format designated as content term diagram was compared with natural language format. Subjects were 48 male undergraduates. Each worked for eight sessions. Results supported interference prediction for time scores and subjects' evaluations of difficulty. However, term diagram summaries resulted in higher information densities and more correct responses to information requests (author).

4/63–73

Bibliographic Coupling between Scientific Papers. 1962. M. M. Kessler. Massachusetts Institute of Technology, Cambridge, Mass. 29 pp. incl. tables, 4 refs. AD–278 831.

The results of automatic processing of a large number of scientific papers according to a rigorously defined criterion of coupling are described. The population of papers under study was ordered into groups that satisfy the stated criterion of interrelation. An examination of the papers that constitute the groups shows a high degree of logical correlation (author).

4/63–74

General Information Manual: Keyword-In-Context (KWIC) Indexing. 1962. Brochure E20–8091 of the IBM Corporation, Data Processing Division, White Plains, New York.

4/63–75

Experimental Study of Convertibility Between Large Technical Indexing Vocabularies. 1962. William Hammond and Staffan Rosenborg. Datatrol Corporation, Silver Spring, Maryland. Technical Report IR-1, prepared under Contract No. NSF C-259.

Every ASTIA Descriptor (in *ASTIA Thesaurus*, First Edition, 1960) and its counterpart in the AEC *Subject Headings* (TID–5001, 3rd revision, January, 1960) were studied. A *Table of Indexing Equivalents* was generated to demonstrate the feasibility of vocabulary convertibility. The written report of this experiment is quite brief — 21 pages. The *Table*, on the other hand, is almost 300 pages. Terms from both lists are arranged by ASTIA Field and Groups. The frequency of use and code number are given for each ASTIA Descriptor. Whenever possible, the equivalent AEC term is listed with a category of equivalence (e.g., 1, identical; 2, synonymous but not identical; 6, AEC term of broader generic level; etc.). Summary tables and a bar graph present statistics on the degree of convertibility between AEC and ASTIA indexing terminologies.

4/63–76

Information Retrieval Study. 1960. John O'Connor. Institute for Cooperative Research, University of Pennsylvania, Philadelphia, Pa. 10 pp. Annual summary rept. no.

1, 1 April 1959–31 March 1960, on contract Nonr-551(35). PB 153 770. Order from LC mi$1.80, ph$1.80.

Contents:

1. The possibilities of document grouping for reducing retrieval storage size and search time.
2. On retrieval in aid of scientific discovery.
3. A note on the possibility of a divided structure file permitting arbitrary substructure searches.
4. Some suggested mechanized indexing investigations which require no machaines.

4/63–77

The Automation of General Semantics. 1960. F. W. Householder, Jr., and J. Lyons. Indiana University, Bloomington, Ind. 11 pp., 2 refs. Quarterly rept. no. 3, 1 Sept.—30 Nov. 1960, on Contract AF 30(602)2184. PB 153 805. Order from LC mi$2.40, ph$3.30.

The subject of the research is the syntactic and semantic analysis of scientific English with the aid of an electronic computer. The primary aim is the development of general mechanical routines for the reduction of complex sentences to their constituent simple sentences without loss of information content. Work so far has been directed to 1) the collection of a representative corpus of scientific writing in English; 2) key-punching and sorting the preliminary more-limited corpus; 3) the elaboration of dictionary lookup and suffix-splitting routines for English; 4) the mechanical determination of the syntactic function of words and of the boundaries of phrases and clauses. Progress in these several directions is described.

4/63–78

The Automation of General Semantics. 1961. F. W. Householder, Jr., and J. Lyons. Indiana University, Bloomington, Ind. 11 pp. Quarterly rept. no. 4. 1 Dec. 1960–28 Feb. 1961, on Contract AF 30(602)2185; continuation of Contract AF 30(602)2184. AD-253 274; PB 155 356 (see also PB 153 805). $1.60.

Contents:

1. Key-punching and sorting of the preliminary corpus.
2. Dictionary lookup.
3. Inflectional and derivational affixes.
4. Walker's rhyming dictionary.

4/63–79

ASTIA Guidelines for Cataloging and Abstracting. 1959. Armed Services Technical Information Agency, Arlington, Va. 22 pp., 3 refs. PB 171 917. $0.75.

These bibliographic procedures are especially tailored for the technical report literature generated by the research and development programs of the Department of Defense. They are not only used by ASTIA for the Department of Defense reports but also with varying degrees of modification by several other report-producing scientific programs both in and out of the government. The purpose of issuing them in this convenient pamphlet form is to encourage standardization of bibliographic techniques for technical reports, thereby accomplishing a threefold result: 1) increasing the effectiveness of the technical report to the scientific community, 2) avoiding unnecessary duplication and reducing the cost of the bibliographic effort, and 3) speeding up the dissemination process (author).

4/63–80

Indexer Consistency Under Minimal Conditions. 1962. J. Jacoby and V. Slamecka. Report RADC-TDR-62-426 of Documentation, Inc., Bethesda, Md.

In the first phase of the project, whose ultimate objective is an improvement of the quality of library indexing, a measure has been assessed of the consistency with which indexers tend to choose the same terms as being descriptive of the same documents. Under artificial indexing conditions which excluded the use of indexing tools, communication, and post-indexing editing, six indexers (three experienced, three beginners) using a Uniterm system of coordinate indexing were found to differ significantly in the number of terms used to index a group of 75 randomly selected chemical patents, as well as in the percent of matched terms any one indexer had with any other. The experienced indexers have attained a significantly higher degree of inter-indexer consistency, with less internal variation, than their inexperienced colleagues. Results of the intra-indexer reliability experiment show that, on the whole, each indexer tends to be consistent with himself when re-indexing "equated" documents and using a vocabulary of "general" (shared) terms; the selection of this vocabulary may have favorably biased these results, however (author).

4/63–81

Translation Title List and Cross Reference Guide. 1961. Edited by F. E. Stratton. Office of Technical Information Extension, Atomic Energy Commission, Oak Ridge, Tennessee. 518 pp. TID–4025(1st Rev.) (Pt. 1). $5.25.

4/63–82

Criteria for Acceptable Abstracts: A Survey of Abstractors' Instructions. 1962. H. Borko and S. Chatman. Technical Memorandum TM–759 of the System Development Corp., Santa Monica, Calif.

The need for criteria by which to judge the adequacy of an abstract is felt most strongly when evaluating machine-produced abstracts. In order to develop a set of criteria, a survey was conducted of the instructions prepared by various scientific publications as a guide to their abstracters in the preparation of copy. One hundred and thirty sets of instructions were analyzed and compared as to their function, content, and form. It was concluded that, while differences in subject matter do not necessarily require different kinds of abstracts, there are significant variations between the informative and the indicative abstracts. A set of criteria for the writing of an acceptable abstract of science literature was derived. The adequacy of these criteria is still to be validated, and the authors' plans for future research in this area were specified (authors).

4/63–83

Automatic Document Classification. 1962. H. Borko and M. D. Bernick. Technical Memorandum TM–771 of the System Development Corp., Santa Monica, Calif.

Starting with a collection of 405 document abstracts dealing with computers, the experiment in automatic document classification proceeds to construct an empirically based mathematically derived classification system by use of a factor analysis technique. The documents are then classified

into these derived categories by five subjects, and the resulting classification serves as a criterion against which the automatic classification is to be evaluated. Of the ninety documents in the Validation Group which contained two or more clue words and which therefore could be automatically classified, 44 documents, or 48.9%, were placed into their correct categories by use of a computer formula. These results are almost identical to the results obtained by Maron in a previous experiment using the same data but with a different set of classification categories and a different computational formula. The experimental evidence supports the conclusion that automatic document classification is possible. Additional experiments are described which, when executed, should improve the accuracy of the automatic classification technique (authors).

4/63–84

Selected Bibliography of the International Geophysical Year: An Example of Tabledex Formats. Robert S. Ledley et al. Joint project by National Biomedical Research Foundation and the Library of Congress. NBR Report 62071/18100. Undated.

The main purpose of this project was to prepare computer programs which would compile a Tabledex Index (a coordinate index with the convenience of a bound-book form). A bibliography of 1,000 articles prepared for the IGY was indexed. The index material was organized and printed by the IBM 7090 computer and consists of six parts: list of retrieval words, alphabetic Tabledex tables, index of authors and special terms, list of retrieval words with frequency code numbers, coded Tabledex tables, and the bibliography proper. The Tabledex tables, the "unique aspect of the volume," are formed as follows: Each table is headed with a retrieval word or frequency-based number. A row of a table consists of retrieval words or number (other than the table heading) that are associated with the article labeling the row. Only those retrieval words are included in a row that alphabetically or numerically follow the table heading.

4/63–85

Coding of Medical Case History Data for Computer Analysis. 1962. Martin Lipkin and Max A. Woodbury. *Commun. ACM*, Vol. 5, No. 10, pp. 532–534 (October).

In this report, items of data conventionally recorded in a medical examination — including history, physical examination, and laboratory tests — were tabulated. The information was then coded to enable automatic transfer to standard punched cards and then to paper or magnetic tape for use in all standard computers. Copies of the form are available to investigators (author).

4/63–86

Syntatic Analysis by Digital Computer. 1962. M. P. Barnett and R. P. Futrelle. *Commun. ACM*, Vol. 5, No. 10, pp. 515–526 (October).

This paper provides an account of the Shadow language that is used to describe syntax and of a corresponding subroutine that enables a computer to perform syntactic analysis. The input to this subroutine consists of a string to be analyzed and a description of the syntax that is to be used. The syntax is expressed in the Shadow language. The output consists of a trace table that expresses the results of the syntactic analysis in a tabular form. Several versions of the subroutine and some associated programs have been in use now for over three years. The present account of the language and the subroutine contains a summary of material that has been described previously in unpublished reports and also some additional discussion of the work in relation to the more general questions of problem-oriented languages and string transformations (authors).

4/63–87

Record Linkage: Making Maximum Use of the Discriminating Power of Identifying Information. 1962. H. B. Newcombe and James M. Kennedy. *Commun. ACM*, Vol. 5, No. 11, pp. 563–566 (November).

Special difficulties are encountered in devising reliable systems for searching and updating any large files of documents that must be identified primarily on the basis of names and other personal particulars. The underlying problem is that of making nearly maximum use of items of identifying information that are individually unreliable but that may collectively be of considerable discriminating power. Rules that can be applied generally to name retrieval systems have been developed in a methodological study of the linkage of vital and health records into family groupings for demographic research purposes. These rules are described, and the ways in which information utilization for matching may be optimized are discussed (authors).

4/63–88

Automatic Indexing: An Experimental Inquiry. 1961. M. E. Maron. *J. ACM*, Vol. 8, No. 3, pp. 404–417 (July).

This inquiry examines a technique for automatically classifying (indexing) documents according to their subject content. The task, in essence, is to have a computing machine read a document and, on the basis of the occurrence of selected clue words, decide to which of many subject categories the document in question belongs. This paper describes the design, execution, and evaluation of a modest experimental study aimed at testing empirically one statistical technique for automatic indexing (author).

4/63–89

Information Retrieval Based Upon Latent Class Analysis. 1962. F. B. Baker. *J. ACM*, Vol. 9, No. 4, pp. 512–521 (October).

Existing document-retrieval systems tend to be based upon the originators' background and skill rather than a mathematical model. The present paper proposed an information-retrieval system based upon latent class analysis. The key words of the documents constituted the basic data in the system. A set of accounting equations based on the joint occurrences of key words were solved for latent class parameters. The analysis then employed the parameters to assign each possible pattern of key words to the m latent classes with a specific probability. The latent classes, however, are a function of the data, not a human librarian.

The storage of information is at two levels: In level one, the possible key word patterns and the corresponding probabilities are stored. In the second level, the documents are stored by latent class. Information is retrieved by using the

pattern of key words in the information request to locate the m probabilities associated with the pattern of key words. The m probabilities are then inspected to locate the largest value, thus indentifying the latent class which has the highest probability of yielding the pattern of key words in the request. Once the latent class is located, those documents in the class having a high probability of belonging to the selected latent class are retrieved.

The salient feature of latent class analysis as a basis for information retrieval is its mathematical basis, which provides a means for further theoretical development (author).

4/63–90

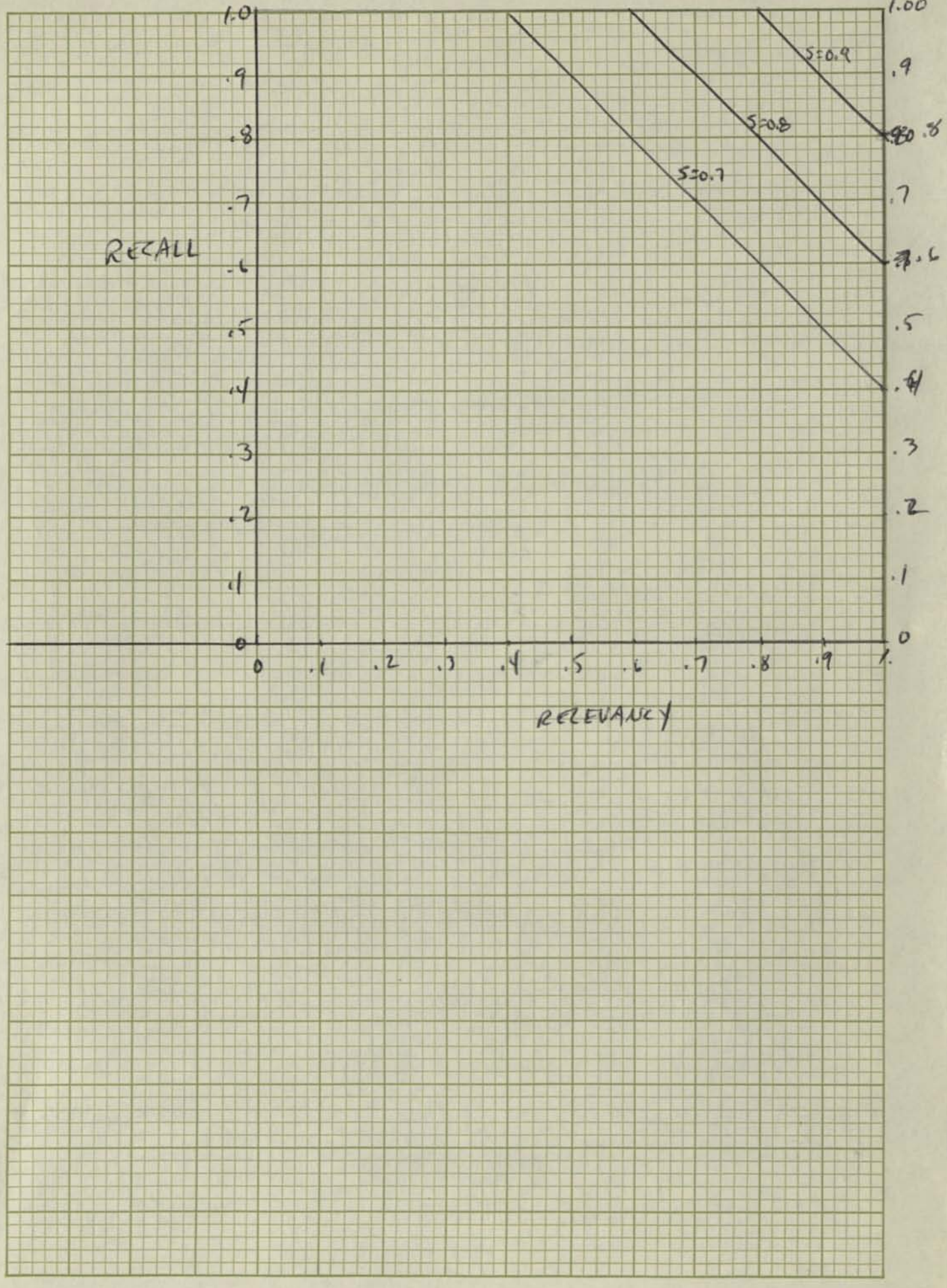
The Power of Words in Technical Communication. 1961. E. P. Odescalchi. *The Technical Writer,* Vol. 1, No. 3, pp. 11–12 (September).

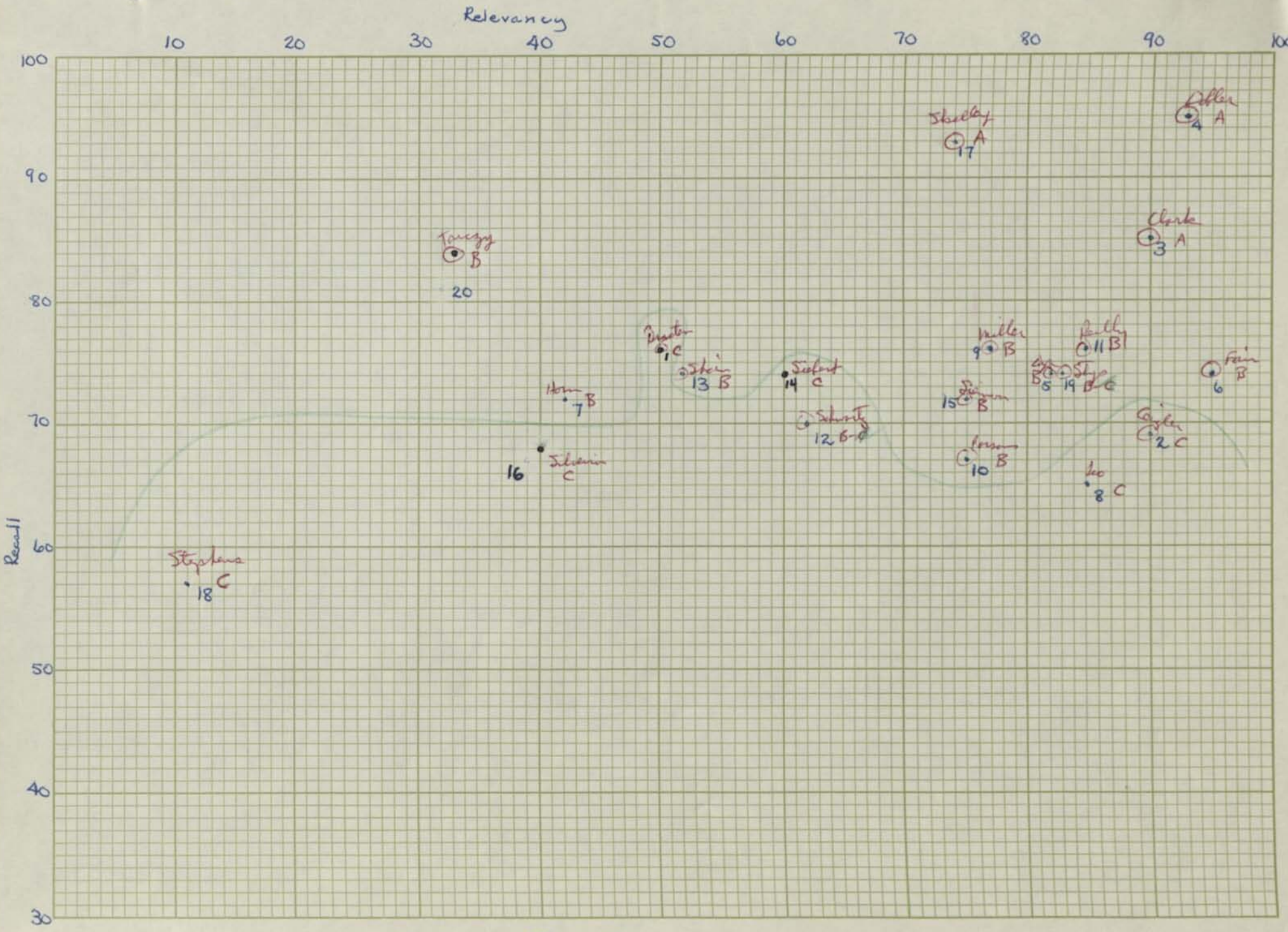
Language is the most common vehicle for transmitting information. The relationship between word power and intelligence is well documented. Intelligence evaluations rely heavily on vocabulary tests, because words are the tools for thinking, and ideas do not exist in the mind apart from them. Statistical data are presented to show the relationship between vocabulary size and degree of achievement in society. People with more words at their command can present their ideas and arguments with greater skill and ultimately win out in the competition with others. All words, even the most complex ones, have very concrete origins. The entire vocabulary of English can theoretically be derived from 461 Indo-European roots. Seventeen master words, mostly Latin, and twenty prefixes which hold the key to about 100,000 words in the unabridged dictionary are compiled, and the former are illustrated with examples (*IBM J.*).

4/63–91

Dissertations in Physics, an Indexed Bibliography of all Doctoral Theses Accepted by American Universities, 1861–1959. 1961. Compiled by M. Lois Marckworth with the assistance of the staff of Advanced Systems Development Division and Research Laboratories, International Business Machines, San Jose, California. xii, 803 pp. NBS Reference Z7141. M3, 1861–1959. Stanford University Press, Stanford, California.

Prepared with the aid of an IBM 704 computer. H. P. Luhn is credited for the KWIK program, modified by Roger M. Simons. There are 8,418 entries in all, and over 5,000 represent degrees granted in 1950–1960. The keyword is brought out to the left margin, and the title is not permuted. There are six information-carrying words/title. The Exclusion List is derived from syntax, roman numerals, units of measurement, and relative terms, such as high, low, etc. Hyphenated words are alphabeted in both places. See and see also references have been added. Machine-alphabetizing sequence is as follows: letters, numbers, and arbitrary symbols. Prefixes were not separated. Numerical keywords were grouped by the measurement expressed and arranged in numerical order as temperature, energy, frequency, and wavelength.





Systems Test Results

Librarianship 245 5-15-63

SYSTEM TEST RESULTS

		RECALL	RELEVANCY
1.	Braaten	.76	.50
2.	Ceizler	.69	.90
3.	Clark	.85	.90
4.	Dobler	.95	.93
5.	Dye	.74	.82
6.	Fair	.74	.95
7.	Horn	.72	.42
8.	Leo	.65	.85
9.	Miller	.76	.77
10.	Parsons	.67	.75
11.	Reilly	.76	.85
12.	Schwartz	.70	.62
13.	Shain	.74	.52
14.	Siebert	.74	.60
15.	Siemon	.72	.75
16.	Silveira	.68	.40
17.	Skelley	.93	.74
18.	Stephens	.57	.11
19.	Stys	.74	.83
20.	Tarczy	.84	.33

Test Committee:

P. F. Clark
I. M. Dobler
G. Skelley
S. Stys

	Inventory Coll. Coverage	Econ. of Micro Storage	Econ. for loan requests	Discovery Time Logs	Inventory Time Logs	Codes & Notations	Inter Library Loan	KWIC	Individual Files	Sample Questions	User Suggestions	
	1	2	3	4	5	6	7	8	9			COMPOSITE
Braaten, Karen	X	OK X	very good X	X	+	-2 -1½ X	X	error late X	needs (missing)	late	(missing)	C
~~Ziegler, Barbara Jean~~			late				X	⊗		X		C
Clark, Prudence (Mrs.)	X	good X	very good X	+	+	OK OK X	X	error X	O	+		A
Doller, Ida (Miss)	+	good X	good X	+	+	OK OK X	X	X	O	+		A
Dye, Sandra	X	OK +	good X	+		-1 -1	X	error X		+	good O	B
Fair, Judy Ann (Mrs.)	X	min. +	OK X	+	+	OK OK X	O	X		X		B
Horn, Barbara	X	good X	good ✓	+	+	-4 -1	✓	⊗	72-42	+		B
~~Kretzmann, Ellen~~												—
Miller, Arnold	X	OK X	OK ✓	+	O	-2 -1	✓	X		+		B
Molin, Leo	X	good X	good X	✓		-2 -1 X	X				(missing)	C
Parsons, Marian (Mrs.)	X	OK +	OK ✓	X	+	OK +1	O	X	O	X		B
~~Pickey, Loretta~~												
Schwartz, Katherine	X	good X	very good X	X	+	OK -1	✓	error X		+	ok O	B
Shain, Chas.	X	very good	very good +	⊗	+	OK +2	✓	error X		+		B
~~Shugart, Carolin (Mrs.)~~												
Silvera, Ronald	X	min. +	OK X	X	+	OK -1	✓	error X	68-40		(poor)	C
Siebert, Roger	⊗	min. late	OK late	X	X	OK -1 X	✓	X	74-60	X		C
Sienov, Fred	+	OK ⊗	good ⊗	✓	incomplete +	OK +6	✓	X		X		B
Shelley, Grant	X			X	+	-2 -1	✓	error X	O	X		A
Styp, Sharie	X	min. ✓	min. X	X	+	OK OK X	✓	error X	O	X		B
Stephans, Mary Louise	+	good ✓	good X	X	+	OK (-1)	✓	X		+	good O	C
Toraczy, Stephan	X	OK +	OK X	+	+	OK (-1)	✓	X		+		B
Reilly, J.	+	good +	good ✓	+	+	OK +1	✓	X		X	good O	B

Students listed below have completed

L. 201, L. 202, and L. 203 and have

indicated they wish to enroll in L.245

Spring 1963

-1-

	Picked up class card on or before February 7, 1963
Braaten, Kaisa (Miss)	
Clark, Prudence (Mrs.)	
Dobler, Ida (Miss)	
Dye, Sandra (Miss)	
Fair, Judy Ann (Mrs.)	
Horn, Barbara (Miss)	
~~Kritzman, Ellen (Miss)~~	
Miller, Arnold (Mr.)	
Parsons, Marcia (Mrs.)	
~~Richey, Loretta~~ (Miss)	x
Schwartz, Katharine (Miss)	x
Shain, Charles (Mr.)	x
~~Shugart, Cornelia~~ (Ann) (Mrs.)	
Siebert, Roger	x
Siemon, Fred (Mr.)	x
Skelley, Grant (Mr.)	x
Stys, Sharie (Miss)	x
Stephens, Mary Louise (Miss)	x
Tárczy, Stephen	x

Mrs. Nyland, Lib. II from Public Health, indicated she would like to audit the course. Needs permission of instructor to do so. - lsg

UNIVERSITY OF CALIFORNIA

Copies for Nve/Wing

SCHOOL OF LIBRARIANSHIP
BERKELEY 4, CALIFORNIA

June 6, 1963

Mr. Charles Bourne
Stanford Research Institute
Stanford University
Stanford, California

Dear Mr. Bourne:

Now that the year is ending let me thank you warmly for your outstanding contribution to our program this spring. The enclosed copy of a letter from Charles Shain should please you. Other students have come to me to speak enthusiastically about your course. I hope that you enjoyed it too and that we will soon have the opportunity to do more of this kind of thing.

Danton and I will be off to Southeast Asia this Sunday, June 9, and will return around the 20th of July. I will hope to see you later in the summer.

Thanks again for an excellent job.

Sincerely,

R. C. Swank
Dean

RCS:hf
Enclosure

Charles Shain
1340 Grove Street
Berkeley 9, California
June 5, 1963

Dr. Raynard Swank
School of Librarianship
University of California
Berkeley 4, California

Dear Dr. Swank:

Please excuse the rather formal device of a letter for communicating with your office by a Library School student (June '63), but even such an inelegant device might serve a useful purpose.

I am writing to you to praise the Spring course in Documentation (L 245), as taught by Mr. Bourne, and to urge that some form of continuation of this course be made available to Library School graduates and/or undergraduates within this area, in the near future.

If you deem it worthwhile and possible, I would suggest two possibilities: (1) A follow-up course by Mr. Bourne, "L 246", or; (2) A special seminar offered under auspices of the Computer Center on the U.C. campus.

Hoping that other students have communicated their privately expressed comments on this matter, I would appreciate hearing about any decision or plans you might reach im this area.

Sincerely,
Charles H Shain