IBM

Reference Manual

Index Organization for Information Retrieval

IBM

Reference Manual

Index Organization for Information Retrieval

TABLE OF CONTENTS

INTRODUCTION

This is a primer on index organization. Only the basic principles are presented and these in a simplified form. There will be no attempt to discuss the problems of subject analysis which the indexer must perform to select the correct index points; rather, the discussion will be limited to the methods and patterns of organizing indexes.

Today, with the development of mechanized information storage and retrieval, there is need for communication between librarians and documentalists on the one hand and systems personnel on the other. It is to help the latter understand the problems he will encounter in organizing information for retrieval that this primer has been prepared.

The literature on indexing is very extensive and its vocabulary is unstable and confusing. The basic principles, however, are not difficult to understand. As the systems man gains understanding of the techniques of information retrieval, he will be in a better position to demonstrate the contributions that mechanization can offer this field.

INDEXING

Indexing is an ordering and listing of names, topics, objects, etc., to facilitate finding the individual items contained in a store of information. The conversion of indexes to codes — that is, the use of special symbols to represent words — is the subject of an IBM pamphlet, Modern Coding Methods (X21-3793). Coding will be touched on only incidentally.

There is no perfect or ideal index organization which is applicable to every situation. Rather, the contents of the file and the uses to which it will be put will determine the form of the index.

Indexing is usually divided into name indexing and subject indexing. Since they serve different purposes and have different patterns of organization, these indexes are nearly always treated separately.

NAME INDEXING

Names are usually arranged in strict alphabetic order, letter by letter, to the end of each word:

Smith, J.
Smith, John
Smith, John A.
Smithell, Alfred

Sometimes it is questionable which part of the name is to be used. The usual practice in the United States is to use the full surname, including compounds, with all prefixes and to file exactly as spelled, disregarding umlauts, accents and other diacritical marks used with foreign names.

d'Alembert	El Al	Macdonald	O'Daniel
Dalton	Fitzgerald	MacRae	O'Keefe
de Secour	Fitz-Hugh	Mayer	Okin
de Vivo	Int'Feld	McCall	Tenant
Devon	L'Abbee	McDonald	Ten Eyck
Disney	LaBelle	M'Lean	Vanner
Di Stefano	Labor	O'Brien	Van Ness
El-Abd	La Chappelle	Obst	Vonner
	MacAllister		Von Rath

Libraries, as a rule, ignore the prefix for foreign names and group the M', Mc and Mac together as if written Mac.

Mace
M'Ewan
MacEwan
Mach
McHale
Macham
MacHatton
McLachlen
Maclay

Indexing of verified names is quite simple. The problem, however, becomes complicated when the exact spelling of the name cannot be established or when a group of people all have the same name. In such instances secondary evidence is introduced to pinpoint the individual. Common items of secondary evidence are birth date, street address, telephone number, Social Security number, signature, physical description such as height, weight, color of eyes, sex, and even fingerprints and photographs.

Where there is doubt about the spelling of a name, the searcher must be able to scan groups of names in order to select the individual he wants. The usual library practice is to cross-reference individual names.

Beam	see also Beem
Behr	see also Baer, Baier, Bair, Baire, Bare, Bayer, Beir, Byer
Beedle	see also Beadle, Beidel
Berch	see also Birch, Burch
Canady	see also Kennedy
Cline	see also Clyne, Klein, Kline
Ebel	see also Able
Eisenberg	see also Isenberg
Lisle	see also Lyle, Lysle
McCloud	see also McLoud, McLeod
McCrea	see also McRea
McElroy	see also McIlroy
Mueller	see also Miller
Philbrick	see also Filbrick
Ray	see also Rea, Wray
Read	see also Reed, Reid
Rhine	see also Ryan
Rogers	see also Rodgers
Saxe	see also Sachs, Sacks
Sinclair	see also Saint Clair, St. Clair
Smith	see also Schmid, Schmidt
Weinberg	see also Wineberg
Ziegler	see also Seigler, Siegler

Cross referencing is sufficient where names are accepted as correct and it is a matter of directing the searcher to the correct entry in the index. Where doubt exists as to exactly what the name is, it may be necessary to have a large number of cross references.

Nickel see also

Niccol	Nichal	Nickell	Nicol	Nikalos
Niccola	Nichala	Nickells	Nicola	Niklas
Niccolai	Nichalas	Nickels	Nicolae	Niklass
Niccolas	Nichali	Nicklas	Nicolais	Nikless
Niccolay	Nichalis	Nicklaus	Nicolas	Nikol
Niccoli	Nichalo	Nickle	Nicolau	Nikola
Niccoll	Nichalos	Nickles	Nicolaus	Nikolaa
Niccolla	Nichals	Nickless	Nicolay	Nikolai
Niccollai	Nicheles	Nickol	Nicoli	Nikolas
Niccollay	Nichels	Nickola	Nicoll	Nikolaus
Niccolls	Nichol	Nickolai	Nicolls	Nikolay
Niccols	Nichola	Nickolas	Nicols	Nikoll
	Nicholas	Nickolay		Nikolls
	Nichole	Nickoll		Nikols
	Nicholes	Nickolls		
	Nicholi	Nickols		
	Nicholis			
	Nicholl			
	Nicholls			
	Nicholo			
	Nicholos			
	Nichols			

Such a large number of cross references, even though they may begin with the same initial letter, are too numerous to be looked up individually. The method usually adopted, therefore, is to group such names under one spelling, treat all variants as if they were identical, and search by the first name. Such a "class" or "bucket" containing all variants can also carry cross references to other classes or single names where the relationship between the names is rather tenuous:

James, Jameson, Jamieson, Jamison see also Jamerson

Phonetic filing is sometimes used to obtain a partial grouping of similar-sounding names. This may involve simply dropping vowels:

Brn	for Braun
Brwn	for Brown, Browne
Jhnsn	for Johnson
Jhnstn	for Johnston, Johnstone

or may involve grouping of similar-sounding consonants. Under one of the more popular schemes:

The initial letter is retained.

W, H are dropped except as initial letters.

A E I O U Y are also dropped but serve as separators.

Remaining consonants are coded up to three figures, as follows:

1. B F P V
2. C G J K Q S X Z
3. D T
4. L
5. M N
6. R

Zeros are added, if necessary, to complete three digits.

Double consonants or equivalents are coded as one letter unless separated by a separator.

Baird	B630
Bird	B630
Byrd	B630
Johnson	J525
Johnsen	J525
Johnston	J523
Johnstone	J523
Johnstown	J523
Jonston	J523

Lowery	L600
Laughrey	L260
Sachs	S220
Sacks	S222
Saxe	S200

As can be seen in the examples, it is not possible to group all similar-sounding names by a phonetic system. Furthermore, special rules must be developed to avoid scattering such similar names as McLane, McClain, M'Lean, or Saint Clair, Sinclair, St. Clair.

Also, a formula approach often groups unrelated or dissimilar names:

Hall	H400
Heil	H400
Hill	H400
Hull	H400
Howell	H400
Howeley	H400

As demonstrated in the "Nickel" example, one must use empirically derived lists of names in order to take care of all possible variants.

There are other techniques for filing names. Although some of these do have the effect of grouping similar-sounding names, their main purpose is to develop short codes, digital representations, or to combine with the name such secondary data as birth date or address in order to develop unique entries. These are coding techniques and are, therefore, not considered here.

ORTHOGRAPHY

So far the discussion has been confined to actual name variants and to variants due to phonetic errors. In some instances where signatures are used, there are errors due to difficulty in interpreting handwriting. In such instances n may be confused with u, r with i, b or h with li, e with i, a with o, and so on. Such orthographic variations can be readily incorporated in a name list.

FORENAMES

Forenames may also be grouped in classes. In fact, this if often necessary because of contractions, nicknames, translations and the like:

> James, Diego, Giacomo, Jaime, Jas., Jim, Jimmie, Vaclav, Venzel, Vincenzo, Waclaw, Wenzel

CORPORATE NAMES

Firm names and other corporate names are treated as personal surnames. Coined names are filed as written:

Backus, J. C. & Company
Belton, Donald F. & William D. Company
Best Brands Inc.
Best, William
Best's Beauty Salon
Bevans and Beverly Service Co.
Beyer, John
Beyer Real Estate
Bill's Barber Shop
Bit of Honey Shoppe
Board of Trade
C & C Auto Service
Commission on Waterways
Committee for Local Government
Consolidated Edison Co.
Cooper Hotel
Co-operative Housing Firm

NOTE: Articles, conjunctions, ampersands, prepositions, etc., are ignored in filing.

At times there is difficulty in determining whether the first part of a firm name should be treated as a forename or used as an entry like a surname:

John Crerar Library
John Hancock Mutual Insurance Co.
John Stewart Methodist Church
Johns Hopkins University
Marshall Field & Co.

The tendency is to file under the first part of the name and to cross-reference from the second part.

NAME FREQUENCIES

The following frequencies, based on samplings by the Social Security Administration, can be of help in setting up name indexes:

Length of Surname

Length in Characters	Percentage	Cumulative Percentages
5 or less	29.53	29.53
6	24.22	53.75
7	21.56	75.31
8	12.81	88.12
9	6.10	94.22
10	2.87	97.09
11	1.15	98.24
12 or more	1.76	100.00

Distribution of Surnames by Initial Letter

Initial Letter	Percent of Total File in Letter	Rank
A	3.051	15
B	9.357	3
C	7.267	5
D	4.783	10
E	1.888	17
F	3.622	13
G	5.103	8
H	7.440	4
I	.387	23
J	2.954	16
K	3.938	12
L	4.664	11
M	9.448	2
N	1.785	18
O	1.436	19
P	4.887	9
Q	.175	25
R	5.257	7
S	10.194	1
T	3.450	14
U	.238	24
V	1.279	20
W	6.287	6
X	.003	26
Y	.555	21
Z	.552	22

The Social Security Administration also publishes a list of some 1,500 most common names arranged alphabetically and by size.

SUBJECT INDEXING

Man has always systematized and organized his knowledge so as better to understand and use it. As the scope of his knowledge has changed and expanded, he has adapted his tools to control it. Today, with the accelerated growth of scientific, technical and commercial information which must be available for use very quickly, and with the development of mechanisms to organize and reproduce large masses of information, there is a crisis in the whole field of information storage and retrieval. Long-established information systems are being reappraised and many new approaches are being tried. The skills and vocabularies of many different disciplines are being brought to bear on the problem. Words are being coined or borrowed from other subject areas to describe the various systems. Thus, although there may be much progress, there is also much confusion.

Much of the confusion can be avoided by relating things to basic principles. In the case of subject indexing there are essentially only three fundamental approaches: classification, subject headings and coordinate or manipulative headings. Practically all specialized indexing systems use one of these approaches or combinations of them. Each has unique qualities and abilities as well as deficiencies. Each must be carefully selected and adapted for the job to be done.

Classification

Classification is a systematic, logical arrangement of index entries usually in a hierarchical or tree pattern. The standard library classification systems, such as Dewey Decimal, Bliss, Cutter, Library of Congress and Universal Decimal, all try to be hierarchical systems. The terms are arranged so that they proceed from the most general to the most specific:

Dewey Decimal Classification

Notation	Term
700	Fine arts
720	Architecture
721	Architectural construction
721.8	Openings and their fittings
721.81	Doors

Library of Congress

Q	Science
QC	Physics
QC 125	Treatises on experimental mechanics
QC 151	Liquids in motion. Hydrodynamics

Highly developed hierarchical systems, such as zoological and botanical classifications, may go through more than 20 steps descending from kingdom through phylum, superclass, class, subclass, infraclass, cohort, order, suborder, family, subfamily, tribe, genus, species, and

so on. Such a logical arrangement of an index is extremely useful. Since it is not necessary to alphabetize the entries, the classified index has the same order in any language, and the language barrier is thus overcome. Class catalogs, therefore, have been very popular in Europe and wherever multilingual groups have had to consult the catalogs and indexes.

Since the position of a topic is fixed and not dependent on language, the synonym problem is eliminated and the need for cross references is reduced. Cross references to show relationships of topics in different classes are, however, necessary and most classification schemes have extensive cross references.

Most important, a hierarchical arrangement permits one to search at any level of indexing. By using an expanding notation, as in the Dewey Decimal system, or some other graded code, the search constraints can be set to include as broad or as narrow a subject as one desires. For example, one wants information on hexose. Depending on the size of the original text and the depth of the indexing used, this information might be indexed variously as:

Hexose
Monosaccharide
Sugar
Carbohydrate

This is actually the hierarchical order, going from the specific to the more general. In an index alphabetically arranged by subject headings, such references would be scattered; in a classified index they would be brought together. A classified index, therefore, employing a code which in its structure reflects the generic relationships of the index, makes for an excellent mechanical retrieval system. It is simple to search at any level of specificity. If a hit is not made at a very specific level, one can automatically go to the next, more general level and so on until a hit is made, assuming, of course, there is informational material on the subject in the file. A classification code number, therefore, not only stands for the input description of a subject in any language, but also brings the subject into some logical relation with other subjects. Further, it provides a simple and efficient address for mechanized storage and retrieval.

Classification, however, has certain disadvantages. An alphabetic index (Dewey calls this the relative index) is needed in order to find where topics are filed:

Topics	Dewey Decimal Classification
Oil	
Animal (chemical analysis)	543
Animal (chemical technology)	665
Baths	542
Burning, locomotives	621
Coal (economic geology)	553
Cooking	641

Topics	Dewey Decimal Classification
Oil (cont.)	
Cookstoves	643
Domestic fuel	644
Feeders (lubrication)	621
Gages (motor vehicles)	629
Heaters	644
Insulating material	621
Lamps	644
Light	644
Motor vehicles	629
Painting (Art)	759
Painting (Building)	698
Plants (Agriculture)	633
Plants (Botany)	581
Refining	614

It is necessary, therefore, to go through two steps to find something. First an alphabetic index must be consulted to find the class number, then the class number looked up to find the reference. This slows the search and makes it more expensive.

Also it is necessary to provide for future expansion of a classification scheme so that new terms may be interpolated anywhere in the scheme. In rapidly developing subjects this can cause difficulty, especially where unforeseen changes occur.

The major difficulty, however, derives from the fact that the demands made on a retrieval system have really nothing to do with logical or hierarchical arrangement. To begin with, there is often no natural basis for a logical arrangement such as is found in biology or chemistry:

Thing
Substance
Chemical compound
Organic compound
Hydroxy compound
Carbohydrate
Sugar
Monosaccharide
Hexose
d-glucose
beta-d-glucose

Rather, most classifications are artificial or synthetic:

Universal Decimal Classification

6	Applied science. Medicine. Technology
66	Chemical technology
669	Metallurgy
669.7	Light metals in general

Universal Decimal Classification (cont.)

669.71	Aluminum. Aluminum alloys
669.713	Extraction of aluminum and aluminum alloys from aluminum compounds
669.713.7	Electrolytic production
669.713.72	Fused salt-bath electrolysis
669.713.723	Electrolysis of aluminum or other oxygen-bearing compounds of aluminum in halide bath

It is really only in nature that one finds a true hierarchy. In almost all other cases it is an artificial or pseudo-hierarchy, sometimes called a chain, representing a particular point of view. There are, therefore, as many workable artificial hierarchies or chains as there are points of view.

In this discussion of classification so far we have used the term hierarchy to describe the relationship between the subdivisions of an index. This is traditional but not very accurate. Actually, all that should be conveyed is that there is a relationship between the topics listed under each index entry. Subdividing a topic does not mean splitting a class into a subclass. Moreover, even where a true hierarchy exists, searching a file need not be hierarchical; in fact, is most likely not to be. For example, if one searcher is interested in dogs as pets, another in dogs as disease vectors, a third in dogs as guardians, none of these searchers derives any benefits from using an index which carefully shows the hierarchical relationships between a specific breed of dogs, canines and mammals in general. In other words, all documents relevant to a given class are not found in that class:

Subject Heading	Library of Congress Classification
Dogs	
Care and breeding	SF427
Diseases	SF991
Folklore	GR720
Legends and stories	QL795.D6
Manners and customs	GT5890
Pictures, illustrations	N7660
Police dogs (Breed)	SF427.S6
Police dogs (Social economy)	HV8025
Taxation	HJ5791
War use	UH100
Zoology	QL737.C2

Recognizing that hierarchy does not meet modern needs, especially of inter-disciplinary literature, a number of people have devised classification schemes in which various classes and categories can be combined at will. A subject file is analyzed to discover the basis for its classification. The various terms are grouped into categories and rules are worked out which govern the order of citation of these categories. Such a classification is often referred to as faceted or "analytico-synthetic." One of the best known systems of this type is the Colon Classification

devised by S. R. Ranganathan. There are also many elements of this free combination in the Semantic Coding developed by J. W. Perry and in the older Universal Decimal Classification scheme. The ability to use separate lists of related concepts, to expand these lists and add to them as needed has made this type of classification a more flexible tool than a classification that tries to be purely hierarchical or, as the colon classifiers call it, "enumerative."

The facet classifiers consider a class a homogeneous subject such as chemistry, physics, medicine, agriculture, history, etc. A category is a differentiation within a class on the basis of various characteristics. In Chemistry, for example, there are categories such as kind, state, property, reaction, operation, device, etc. Alcohol is a kind of chemical, liquid is a state, volatility is a property, combustion is a reaction, analysis is an operation, and a flask is a device. In the class Medicine there are such categories as organs (heart), problem (disease), symptom (fever), agent (virus), handling (surgery), etc. Within the categories there can, of course, be hierarchies.

The order in which these categories are to be arranged can be prescribed so that, for example, an organ is always first, a problem is second, a symptom third, a handling fourth, and so on. Thus an article describing the use of penicillin to cure an inflammation of the skin would read

Skin - Inflammation - Therapy - Penicillin

Using a proposed faceted classification for nuclear energy, the notation

R212. $2D_2O$-081. 2-071AIR-061-022

means

"Start-up of thermal reactor, moderated by D_2O using enriched uranium fuel with air coolant, for research."

R2 = Reactors
R212. 2 = Thermal reactors
D_2O = (Heavy water)
081. 2 = Enriched uranium (used as fuel in a reactor)
071 = Gas cooled
AIR
061 = Research
022 = Start-up

The facets in this example are linked by dashes. Other linkages and relationships can be shown by using colons, zeros, or apostrophes. Using examples of the Universal Decimal System:

538.114:669.245.3 = Ferromagnetism of nickel copper alloys

538 = Magnetism
538.114 = Special theory of ferromagnetism
669 = Metallurgy
669.2 = Nonferrous metals
669.245 = Nickel alloys
669.245.3 = Copper-nickel alloys

621.365.2.078 = Automatic regulation of arc furnace
546.623'32'226 = Potassium aluminum sulphate

An example of another faceted classification is:

CcIufNbj = Transonic flow over a bent airfoil
Cc = Airfoil
Iuf = bent
Nbj = transonic flow

A colon classification example would look like this:

L2153:4725:63129:B28 = Soft palate - Cancer - Radium Treatment - Statistical study

L = Medicine
L2 = Digestive system
L21 = Mouth
L215 = Palate
L2153 = Soft Palate
L2153:4 = Disease and so on

An example of the Semantic Code is:

MWTL.PASS.RQHT.001 = Heat treating

MWTL = Metal
PASS = Processing
RQHT.001 = By means of heat

Nevertheless, such synthetic or artificial classifications, when developed, still represent, individually, a single rigid approach to a subject. A fixed classification, as has been shown, often does not coincide with the needs and viewpoint of the searcher, nor does it really avoid the problems of expansion. This does not mean that classification is not a valuable tool in the preparation of indexes. Under certain circumstances it makes for a good index and it can also be helpful, as will be shown, in the preparation of alphabetic subject headings.

Classification, in general, is better suited for well-established subjects where there is not much change or expansion. And it is better suited where the index users have a single, unified and rather specialized viewpoint. If a library is concerned with basically a single subject and the users of the index or catalog have either a uniform viewpoint of the subject

matter or at least understand or are in agreement as to the organization of that subject, then a classification scheme can be very useful.

Subject Headings

Most American libraries use a classification scheme to arrange books and other publications on their shelves but use alphabetic subject headings to catalog and index the collection. An alphabetic subject index uses a single word, phrase or noun combination that fully and exactly identifies the subject matter:

Astatine
Civil engineering
Flower arrangement, Chinese [Japanese, etc.]
Gases - Liquefaction
Ionization in water
Ionization of gases
Maps, Military - History
Mathematics as a profession
Packaging - Materials, Aluminum
Shielding (Electricity)
Shielding (Radiation)
Heart - Diseases - Research
Tungsten - Physical properties - Tensile strength - High temperature
Uranium - Rolling (Alpha-phase)

An alphabetic subject index is an extremely efficient tool for finding specific subjects. It has only one arrangement and is self-indexing. Access to each subject is direct. Natural language is used and no transformation into a class or code is necessary. The public can use it without special instruction. New terms may be introduced whenever and wherever needed.

The main problem with subject headings is to bring the vocabularies of both the index and the searcher into coincidence, so that the information sought is not missed. In other words, the searcher coming to the index must use the same words in the same order as the index does, in order to find the entries he is seeking. Generally speaking, language has a fairly stable semantic history, and many names of elements, materials, concepts and forms are unique and fixed. The same terms are used in many different indexes over long periods of time. In some subjects, such as chemistry, the terms used are often generated by accepted rules and are unambiguous.

There are, on the other hand, many synonyms, near synonyms, overlapping terms, vague terms, erroneous and superseded terms and other possible sources of terminological difficulties. Most of these can be overcome by providing adequate cross references of the "see" and "see also" variety:

Airstrips	see Airports - Runways
Berlin airlift	see Berlin - Blockade, 1948-1949
Boring machinery	see also Rock drills
Distillation apparatus	see also Column packing; Evaporators; Packed columns
Invertebrates	see also Arachnida; Anthropoda; Brachiopoda; Coelenterata; Crustacea; Echinodermata; Insects; Larvae - Invertebrates; Mesozoa; Mollusks; Myriapoda; Polyzoa; Protozoa; Sponges; Worms
Medical care plans	see Insurance, Health; State medicine
Medical examiners	see Coroners and medical examiners

Some cross references are more elaborate and even resemble thesauri:

Counting devices	Electrical or mechanical devices for registering or recording numbers, not to be confused with radiation detection instruments which are often called counters see also Radiation detection instruments; Radiation detectors; Scalers
Heart - Diseases	see also Angina pectoris; Arrhythmia; Chest - Diseases; Coronary heart disease; Endocarditis; Heart - Valves - Diseases; Rheumatic heart disease
Indians - Legal status, laws, etc.	see also subdivision Legal status, laws, etc., under names of groups of Indians and names of individual Indian tribes; e.g., Indians of North America - Legal status, laws, etc.; Cherokee Indians - Legal status, laws, etc.
Mental health laws	Here are entered works on laws dealing with the care of the insane, the mentally ill, the mentally handicapped, alcoholics, epileptics, and narcotic addicts. Works dealing separately with alcoholics, epileptics, or narcotic addicts are entered under the specific headings. Works on the legal status of the insane are entered under the heading Insanity - Jurisprudence.

Such explanations, usually referred to as scope notes, are effective not only in defining subject headings but also showing exactly the categories in which they fall and their range of applicability.

The problem is somewhat more complicated where terms for new concepts must be chosen. In the areas where language has not been stabilized, the choice of the correct term may have to be tentative and subject to later revision. This, however, is easier to do than to try to find a new slot in a classification scheme.

Another source of language difficulty is the tendency for information requesters not to formulate their questions precisely. Generally speaking, they tend to phrase their inquiries in the broadest terms, asking, for example, for a treatise on physics when they really want to know the slow neutron cross section of zirconium. To overcome this, librarians build a pyramid of cross references going from the general to the specific and making cross references to related subjects:

Engineering	see also Civil engineering
Civil engineering	see also Mining engineering
Mining engineering	see also Petroleum engineering
Petroleum engineering	see also Oil wells

Since classification provides at least one hierarchy, the need for such cross references is somewhat reduced in classification schemes, but is by no means eliminated.

In addition to cross references, sometimes multiple entries are provided for the various related terms so that no matter where a searcher enters the file he will find the desired references. Multiple entries, however, can be used only very sparingly; otherwise the index will become too large to handle.

Particles	see also headings such as Nickel powders see also Alpha particles; Beta particles; Charged particles; Dusts; Elementary particles; Nuclear particles; Powders; S particles; T particles; V particles
Charged particles	see also Ions; Particles
Dusts	see also Aerosols; Particles; Powders
Elementary particles	see also specific particles, e.g., Mesons and V particles. For elementary particles with zero spin, see also Bosons and for those with nonintegral spin see also Fermions see also Antiparticles; Strange particles
Nuclear particles	see also the specific particles concerned see also Elementary particles; Nucleons; Radiation
Powders	see also powders of specific elements see also general headings of the form Oxide powders in the list below for lists of powders of specific compounds see also Fluoride powders; Glass powders; Graphite powders; Hydride powders; Metal powders; Oxide powders; Particles; Steel powders; Sulfate powders; Sulfide powders

Another approach is to group terms into small classifications so as to bring like things together. In order to preserve the alphabetic order of the entries, the usual technique is to invert the subject heading and thus make the noun the file word:

Geometry, Algebraic
Geometry, Analytic
Geometry, Descriptive
Geometry, Differential
Geometry, Enumerative
Geometry, Infinitesimal
Geometry, Plane
Geometry, Projective
Geometry, Solid

Some alphabetic subject heading indexes tend, therefore, to be hybrid schemes, for they include small class groups in what are otherwise direct entry lists. Modern research libraries, however, prefer not to use inverted headings and, instead of class groupings, rely on cross references.

In order to make logically connecting cross references and thus tighten the connective structure, indexers and catalogers sometimes first develop classified chains of hierarchical definitions. Such a systematic classified list is then used to develop the actual subject headings and their scope notes, which define them, in order that the headings be precise and not overlap. In other words, a classification can be a guide for the development of subject headings and cross references.

For example, the hierarchy or "chain" shown on page 11:

Organic compound
Hydroxy compound
Carbohydrate
Sugar
Monosaccharide
Hexose
d-glucose
beta-d-glucose

tells the indexer that cross references from any one of these terms should be made to the others. But, as was explained in the Classification section, there can be several different hierarchies for Sugar, for example, and therefore this chain is only partially helpful in making cross references.

Since compound subject headings are usually required to describe adequately an entry, the possible permutation of terms can cause difficulty. Entries might appear variously as:

Copper-tungsten-zinc alloy - Phase diagram
Zinc-copper-tungsten alloy - Phase diagram
Tungsten-zinc-copper alloy - Phase diagram
Alloys - Copper-zinc-tungsten - Phase diagram
Phase diagrams - Copper-zinc-tungsten alloy

This problem has never been adequately solved. A few conventions such as listing the constituents of alloys, cermets, etc., in alphabetic order as in the first example can help a little. General vague rules such as putting the "most significant" word first, or developing categories of words — realization, material, processes and problems, place, time, form — and assigning an order to these categories, as do the facet classifiers (see page 13) really do not help very much. Very detailed indexes permute or "rotate" the entry word and so provide multiple entries rather than use "see also" references. In general, however, such a multiplicity of entries will bulk a manual index so that it becomes difficult to use.

Although subject headings can be very precise, from a practical point of view they are usually not as precise or detailed as they should be. This is due to the fact that the indexer or cataloger, for reasons of economy, usually indexes to the level of the document rather than to the level of the concepts in the document. For example: Two documents are received, one a brief account on the tensile strength of zirconium at 800° F, the other a large report with very elaborate tables and graphs giving all the known physical properties of zirconium. The first document would be indexed:

Zirconium - Physical properties - Tensile strength - High Temperature

The second document, which actually has much more detailed information on the high temperature tensile strength of zirconium, would be simply indexed as:

Zirconium - Physical properties

The unsophisticated searcher coming to the index or catalog looking for the high temperature strength of zirconium would find the first document but not the second, unless he took the trouble to read through all the entries under the broader headings. Conversely, anyone approaching the index by the broader heading Physical properties might miss the first document.

Librarians have, of course, prepared separate index entries for various portions of a book. Such "analytics" have been used primarily where a publication covers a variety of topics that cannot be grouped conveniently. Analytics have also been used to bring out subjects for which the library does not have separate publications.

Indexers sometimes use broader headings and rely on the bibliographic information carried with the entry to help the searcher select the specific references he needs. On unit library catalog cards, the full author and title and often an abstract or notes give a great deal of specific information not covered by the subject heading. In indexes of abstract journals, unless the complete bibliographic entry is included under each subject heading (Index Medicus), the usual practice (Chemical Abstracts) is to have a descriptive phrase with each entry.

Unit Catalog Card

Welds - Tensile properties
Battelle Memorial Inst., Columbus, Ohio
Causes of cracking in high-strength weld metals, by A. J. Jacobs, R. P. Sopher and P. J. Rieppel. Report on Contract AF 33(038)12619. August 54, 35p. 5 refs.
WADC Technical Report 52-322, Pt. 3; AD-65 474 PB 145 332

Hot-tension and weld-metal cracking studies were conducted on SAE 43XX-type steels and other selected steels. Results from these studies showed a correlation, inasmuch as an increase in carbon, sulfur, and phosphorus tended to lower hot ductibility and promote hot-crack susceptibility.

Bibliographic Entry

Chromosomes - Metabolism
Lima de Faria, A.
Incorporation of tritiated thymidine into meiotic chromosomes.
Science 130:503-4, 28 Aug 59

Descriptive Phrase

Stratosphere
fall-out, transport and mixing, 14:9306
Sulfur dioxide
absorption and diffusion in basic Al sulfate solns., 17341f

In mechanical retrieval systems, until very recently, it has been impossible or certainly uneconomic to store extensive bibliographic and descriptive information along with the entry. This technique has, therefore, not been used and greater reliance has been put on multiple subject headings.

In modern scientific and technical research, much of the information retrieval consists of searching for precise data. The indexes, therefore, are becoming more and more detailed, to the point that some indexes are larger than the body of information they index. The ideal complete index is, of course, a concordance, in which practically every important word is indexed; this is only rarely practical. Also, since the rate of publication is rapidly expanding and the various subject bibliographies, abstract journals and other bibliographic tools are becoming more comprehensive in their coverage, such detailed indexes are becoming too large to be properly searched by manual methods.

Indexes are, therefore, growing much faster than even the rapid growth of literature itself. The information sought is extremely detailed and the index must provide for every level, from the most specific to the most general, and must provide for every possible approach that the inquirer might choose. Classification schemes and subject headings are essentially based on past experience. It is impossible for the indexer to predict the

viewpoint of a future inquirer. With the headings fixed, it is often impossible to extract new concepts which may be contained in the recorded information.

Coordinate Indexing

With the development of punched cards, both hand-sorted and machine-sorted, information has been recorded in fields on the cards and then the cards have been searched by combining various fields to extract the information sought. The standard example is the payroll-personnel record where discrete fields are set aside for age, sex, salary, location, skill and the like. These fields can then be combined at will and searched to find certain individuals with certain qualifications. This system of combining terms is usually called coordinate indexing, but has also been referred to as manipulative indexing, post combination indexing, multi-aspect indexing, multi-dimensional indexing, etc.

Similarly, the individual terms which make up a subject heading can be coordinated or combined at will at the time the search is made. These terms are variously referred to as descriptors, keywords, key terms, discriminators, identifiers, or Uniterms. For example, using a subject heading mentioned in a previous section,

Zirconium - Physical properties - Tensile strength - High temperature

a card would be prepared for each descriptor used (term card) and the document numbers, referring to the documents that contain this information, punched into these cards. When references are wanted covering this complex subject, the appropriate term cards are pulled and matched. All document numbers which appear on all four cards will contain information on the high temperature strength of zirconium. If one is searching for the more general topic of the physical properties of zirconium, then a match of the two term cards Zirconium and Physical properties will also retrieve these documents.

This coordination of terms removes all the need for permutation, since order of terms makes no difference. It also enables searching at any level of specificity without the need for multiple indexing.

The fact that coordinate indexing generally includes all the specific entries in the general heading causes some difficulty. For example, entries for all specific breeds of dogs will also appear on the term card Dogs. This means that when one is searching for general information on dogs, one will get all information in the files on dogs including all individual breeds, everything on their therapy, training, history and so on. This means that general topics are so overwhelmed with specifics that the former are not useful as searching points. To overcome this, the indexer usually employs the descriptor General to segregate general works on a topic. In other words, a general book on dogs would be indexed on the Dog card and the General card. By combining these two descriptors, this book would be separated from all the specific texts on dogs.

The matching of terms to find information is not efficient in manual systems but lends itself well to mechanized and semi-mechanized procedures. In manual systems, the term cards must be pulled from the file and refiled. This cannot be left to the public. The actual visual matching of numbers is a fatiguing process. Searching too is "blind" in that there is no bibliographic information with each entry to assist the searcher in making a selection. There are also other problems involving posting to update the files.

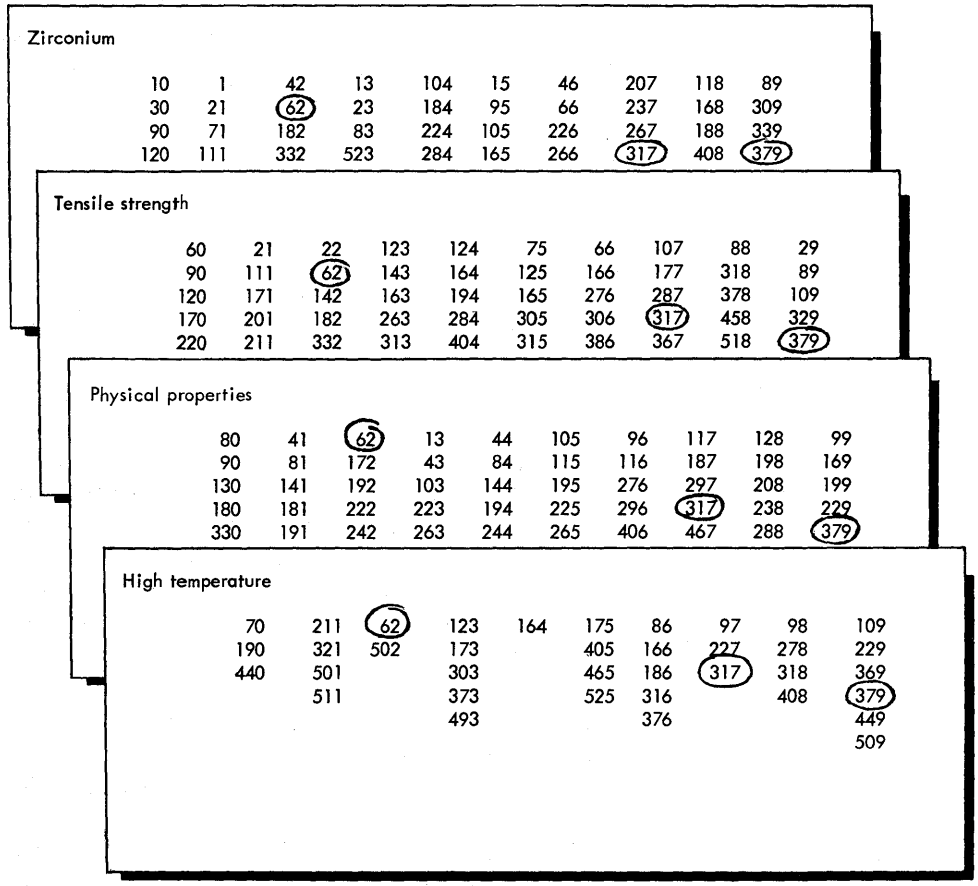
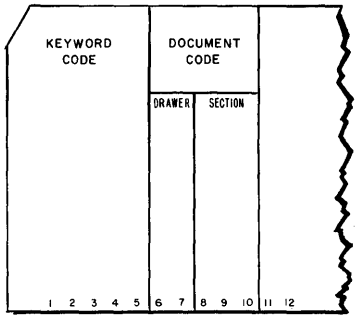
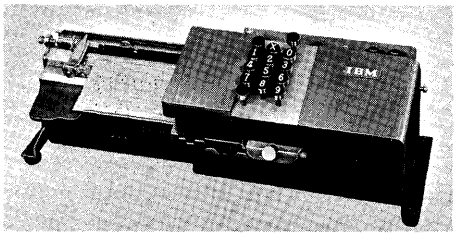


Figure 1. Uniterm Cards - Manual matching

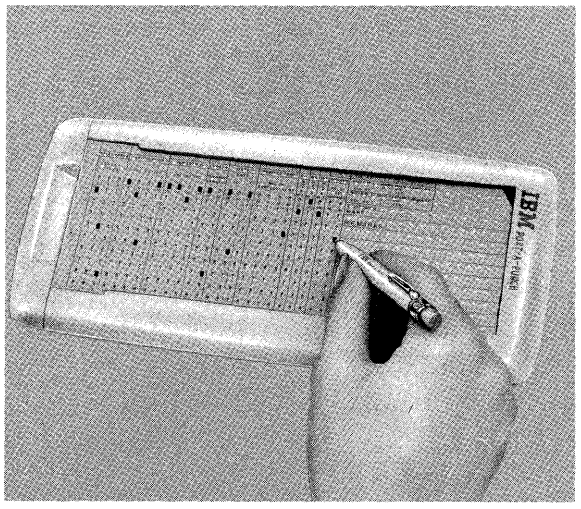
In mechanized systems, however, matching of numbers or holes in a card can be done efficiently, quickly and accurately. Coordinate indexing, therefore, has become popular for mechanized retrieval. At the simplest level it is just a visual matching of holes in punched cards. Such a system involves setting up a card for each term and filing the cards in alphabetic order. As the documents are received, they are numbered and all the descriptors applicable to a document are recorded. The cards carrying these descriptors are pulled and the position which bears the identification number of the document is punched. This can be done manually by removing the chips from a prescored card with either a pencil or a simple Port-A-Punch®. The cards are refiled and the process repeated for all subsequent documents.



The index cards can be punched with an IBM 24 Card Punch, an IBM 10 Card Punch or an IBM Port-A-Punch.



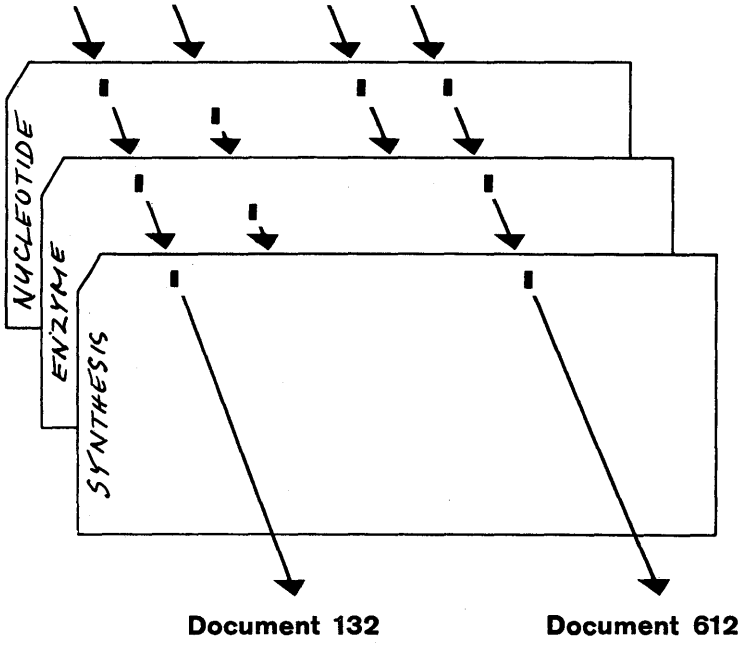
IBM 10 Card Punch



IBM Port-A-Punch

Figure 2

To search the file, the key term cards which characterize the information sought are pulled. The cards are stacked with their edges evenly aligned. The stack of cards is held up to the light. Where holes coincide, light will come through. These will represent the document numbers sought. This simple coincidence of holes is referred to as the Batten or peek-a-boo system.

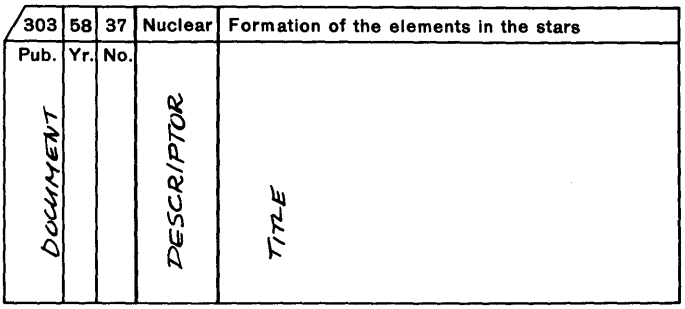


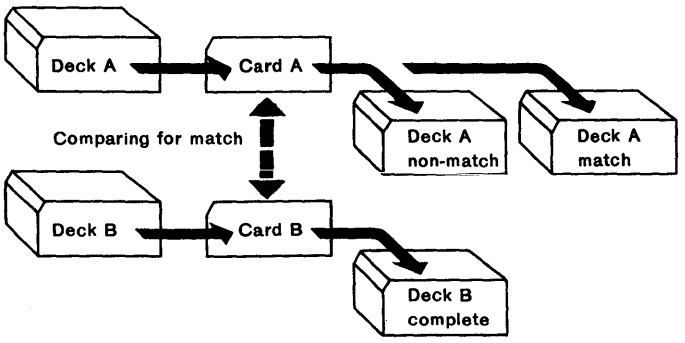
Position coding of document numbers. Where the beam from a light source shines through the selected cards, the hole represents a document indexed under the descriptors stated in query.

Lookup with the IBM card "peek-a-boo" method

Figure 3

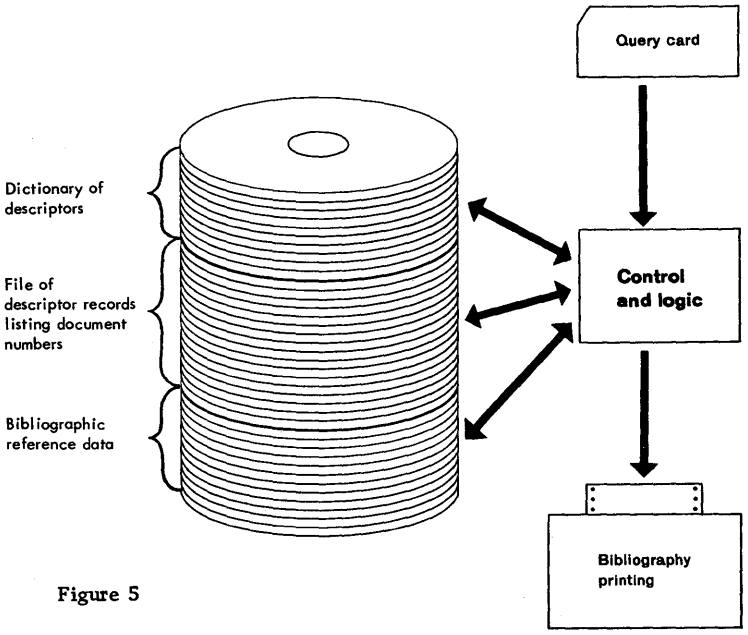
In mechanized systems one of two basic approaches is used, depending on whether the index is searched serially, or whether the entries are prefiled by arranging the items under each term. In a prefiled system, a unit card (term card) is prepared for each entry. Coded into the card are the document number and a term. There are as many cards made for each document as there are terms used to index the document. Term decks are kept separately in document-number sequence. Whenever a subject is searched, the appropriate terms decks are selected and matched with a collator. A similar matching can be done with entries stored in a RAMAC® system.





Lookup with the IBM Collator

Figure 4

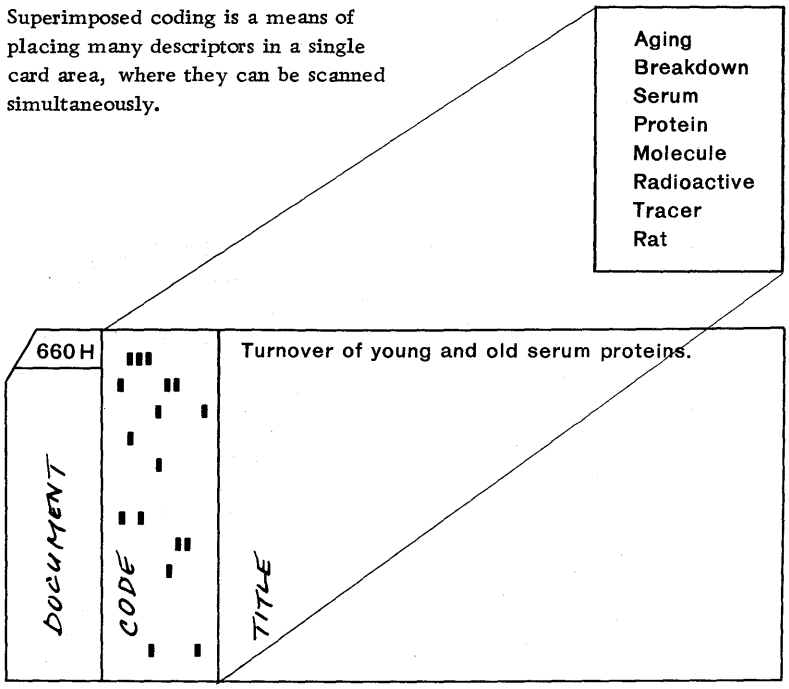


Lookup with a RAMAC system for Information storage

Capacity of many millions of characters

Figure 5

In a serial search system, a card is prepared for each document (item card). On the card is coded the document number and all the descriptors applied to the document. In conducting a search using an IBM 101 Electronic Statistical Machine, the control panel is wired to compare for the presence of individual descriptors. Those cards which have all the descriptors sought are segregated into one pocket, or their identification numbers are printed out or duplicated on other cards. Since, however, the search question may have too many terms and thus reject useful references, subsearches can be carried on at the same time. The machine can, therefore, also segregate all cards which meet all requirements but one, all requirements but two, and so on.

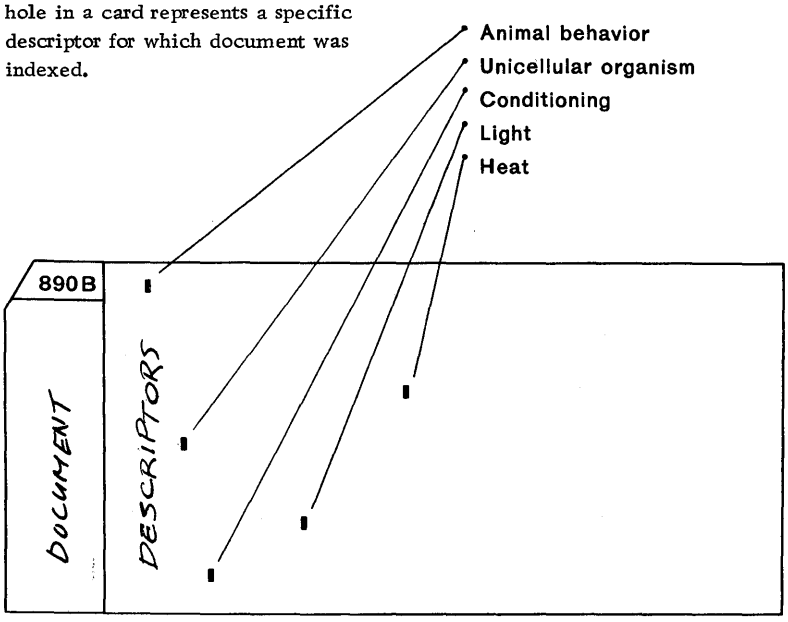


Search with the
IBM 101 Electronic Statistical Machine

Figure 6

The same technique can be applied using the 108 Card Proving Machine, the 101 with the row-by-row scanning device, the Universal Card Scanner, or any magnetic tape system. It can even be used with an ordinary sorter. To increase the speed of selection with the sorter it is advisable to use the Multiple-Column Selector feature and to use a single punch to code each descriptor.

Position coding of descriptors. Each hole in a card represents a specific descriptor for which document was indexed.



Search with the IBM Sorter

Figure 7

Although Figures 8 through 10 illustrate cards and files used with the Universal Card Scanner, the same patterns can be used with any serial searching machine.

Figure 11 illustrates the preparation of a dictionary or authority list of the terms used in the index. Although in this case this dictionary is used to control the assignment of codes, it is also used to control the assignment of descriptors as shown in Figure 12 so that there will be uniformity of terminology and thus no scattering of information.

As noted, coordinate indexing avoids the need for permutation completely. It makes no difference in what order the descriptors of a complex subject heading are arranged. All the documents containing information

Copper-tungsten-zinc alloy - Phase diagram

will be found if one approaches the coordinate index by copper, tungsten, zinc alloy or phase diagram.

The search parameters can be set at will. All the documents found in the above example will also turn up if only phase diagram is searched, if alloy is searched, if copper and alloy are searched. In other words, no document will be missed, no matter what combination of terms is used. The more terms combined, the greater the search constraints. The fewer terms used, the broader the search.

There are three major difficulties, however, with coordinate indexing, and special techniques must be adopted to minimize them. These are false coordination, incomplete coordination and the necessity to show relationships.

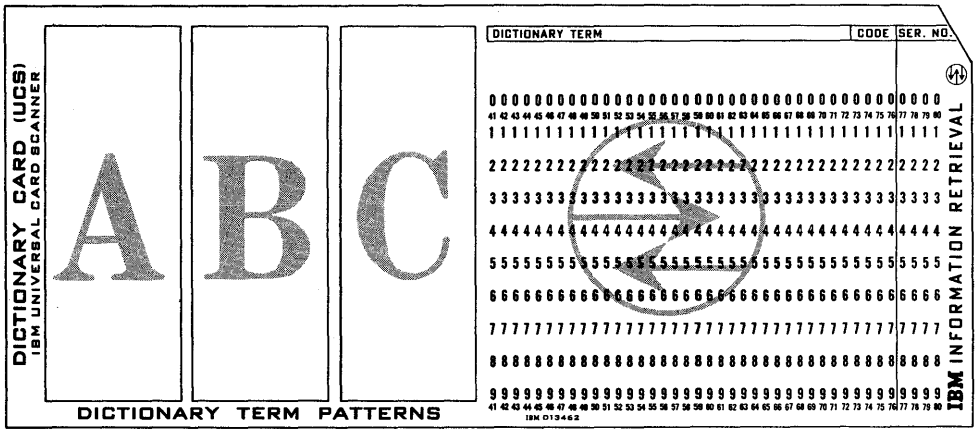


Figure 8a. Dictionary Card, Front

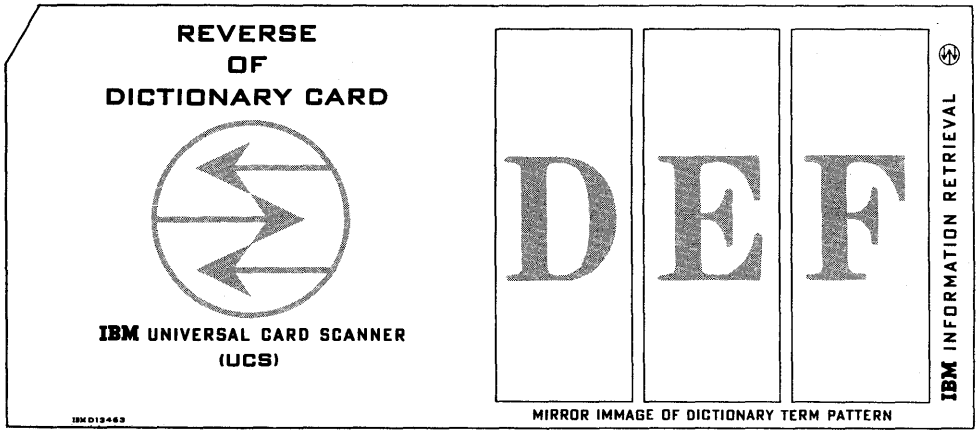


Figure 8b. Dictionary Card, Reverse

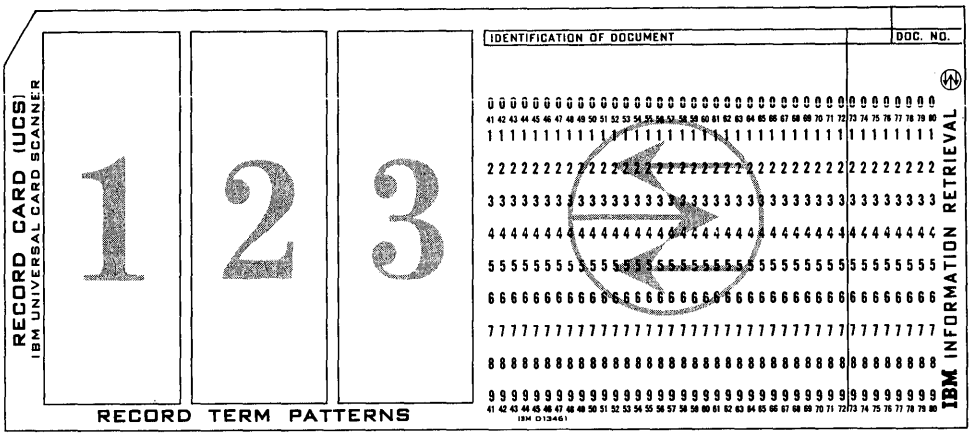


Figure 9. Record Card

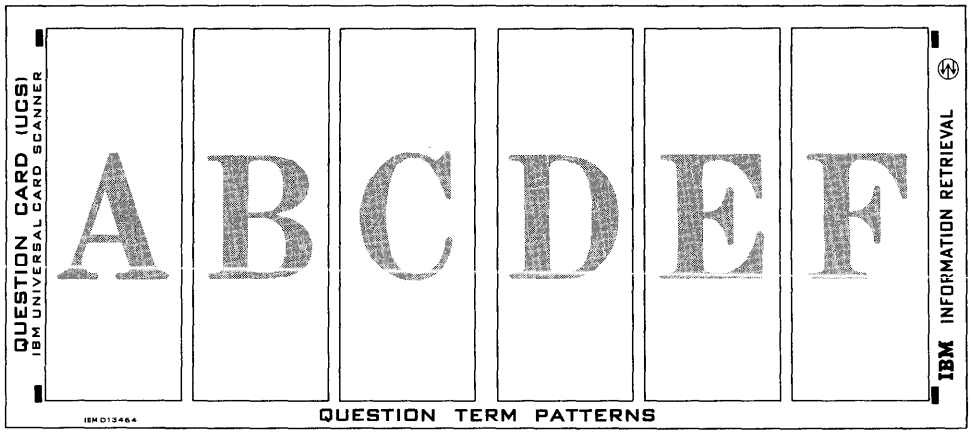


Figure 10. Question Card

PREPARATION OF RECORD CARD

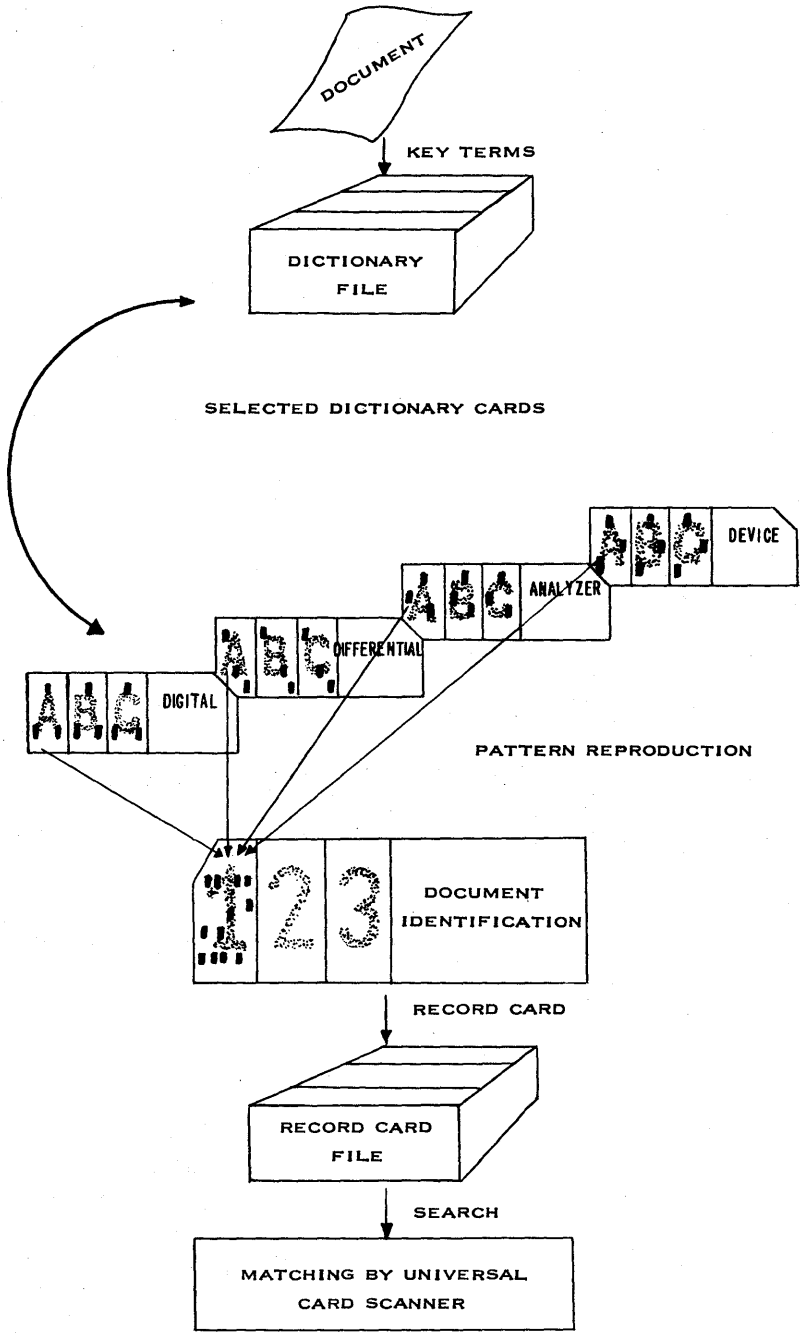


Figure 11

PREPARATION OF QUESTION CARD

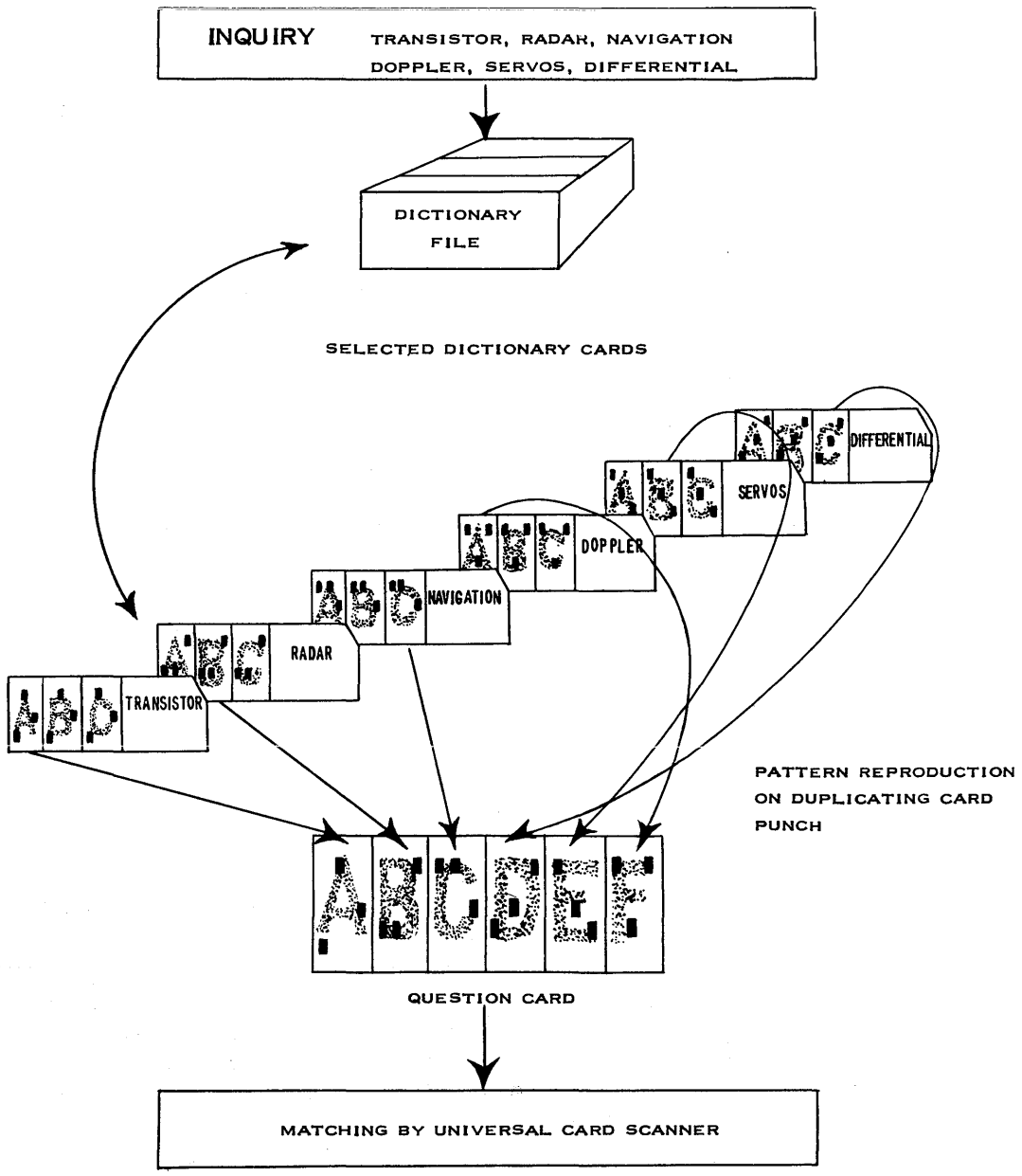


Figure 12

FALSE COORDINATION

If a document contains a series of complex subjects:

A and B and C
also D and E and F
also A and C and F

a search for subjects AEF and DBC will produce this document. This is a "false drop," or false selection, since there is no information on AEF or DBC in this document. Since the descriptors A, B, C, D, E and F all refer to the same document number, they will all match during a search and false drops will occur. One solution is to segment the document and number each section separately, assuming, of course, that each indexable subject is in a separate section. This is often not practicable. Another approach is to apply a symbol to each document number associated with a term and only the document numbers which bear the same symbol can be coordinated. In the above example, the first subject might use symbol 1, the second symbol 2 and the third symbol 3. This document number would, therefore, carry the symbol 1 on term card B, symbol 2 on term cards D and E, symbols 1 and 3 on term cards A and C, and symbols 2 and 3 on term card F. Such symbols have been referred to as "interfixes," "modulants," "role indicators," and "association links."

INCOMPLETE COORDINATION

In the above example of a copper-tungsten-zinc alloy, this reference will be found when searching for copper-zinc alloys, copper-tungsten alloys and tungsten-zinc alloys. This is an incomplete coordination since the search constraints fall within more complex subjects — that is to say, a copper-zinc alloy is quite different from a copper-tungsten-zinc alloy. Generally speaking, where this is a problem — that is, where a topic cannot be broken — it is necessary to use "bound terms" (meaning that the individual descriptors cannot be separated), which is really using a subject heading instead of descriptors. Radio frequency might be considered a bound term which has to be distinguished from Radio and Frequency as separate terms. Also in the illustrations for Uniterm cards (Figure 1), Physical properties and Tensile strength are shown as bound terms. Where bound terms are used, the benefits of coordination are lost. In many instances, however, incomplete coordination occurs very seldom and a few false drops are tolerated.

NEED TO SHOW RELATIONSHIP

For some types of information, the mere juxtaposition of terms is sufficient to describe the subject. There is no ambiguity about

Aluminum - Hardness tests
Cancer - Therapy

but what do the following mean?

Paint - Bacteria - Effect
Uranium - Analysis
Paper - Pulp - Preparation

Is this the effect of paint on bacteria or bacteria on paint? Is this an analysis of uranium or for uranium? Is paper being prepared from pulp or is pulp being prepared from paper? Is a term a subject, object or modifier? In other words, relationship between terms, or the syntactic role of terms, is often very important. These relationships can be temporal, spatial, kinetic or logical. They can show the relationship between specific and generic, between starting and final material, between parasite and host, part and assembled complex; it might involve direction of action, etc. In patent searching, for example, it is necessary to distinguish between the process, the apparatus, the product, the starting material, the intermediate product, the end product, and so on. Such relationships, usually expressed by prepositions and verbs, are normally lost in coordinate indexing, but they can be expressed by adding symbols, modulants, interfixes or role indicators. The particular relationship can be denoted either by particular symbols, or by the joint presence of two or more symbols, or by the order of the symbols.

As an example of using a particular symbol, the addition of symbol 1 on a term (a name of a drug) means that this is a pretreatment drug and is not the actual physical agent.

As an example of using the joint presence of two symbols, the subject could be the preparation of silicon tetrachloride from silicon. Symbol 5 applied to silicon tetrachloride means that this is the entity prepared, fabricated or analyzed for; symbol 1 applied to silicon means this is the raw material.

As an example of showing relationship by order of symbol, if a term is coded in the first field, it means it is the chemical under test, but if it is coded, say, in the second field, it means it is just a chemical used in the process.

One should not exaggerate the importance of showing relationships. In many instances it is either not necessary or the meaning is unambiguous. Some systems insist that a role indicator be applied to every term, so that, for example, a term like Telephone is so constructed that the basic generic relationships of this word — namely, Device, Transmission, Information, Electricity — are all indicated. It is extremely doubtful that such relationships would ever be sought in an index. Furthermore, a few simple cross references could take care of all the normal generic relationships in this instance. (For further discussion of role indicators see section Indicative and Informative Indexes.)

SPECIAL INDEXES

Word Indexing and Subject Indexing

Word indexing uses words as found in the material and indexes them with a minimum regard for standardized meaning. This is a form of indexing which has been practiced widely for over 100 years by European libraries and involves the use of certain words from the titles as entries for a catalog. Recently it has been proposed that all the important words on every page of the documents to be indexed be marked and used as index terms. This high-density type of indexing — as many as 50 terms per page — would ensure that no information be lost. The original proposals for coordinate indexing also were based on the concept that actual words of the text would suffice as the descriptors.

The difficulty is that word or title word entries are inconsistent. Different names are used by different authors for the same subject. Synonyms, author inconsistencies and metaphors will scatter entries throughout the alphabet and no amount of cross referencing can bring the like subjects together. An English writer will speak of maize, valves and wireless, whereas an American author will use corn, tubes and radio. A farmer will speak of wheat and barley; a botanist will use triticum and hardeum. "The light that failed" is not about lights but about eyesight. One man will say heredity, another inheritance and both mean the same thing. And so on.

Word indexing works well for indexing a single work of a single author. It even works for a relatively small group of publications in a limited subject area. It breaks down, however, when applied to any large collection or a variety of subjects.

Subject indexing really involves subject analysis of a document and the selection of the significant standardized terms to describe the contents. The significant information in the document may be expressed or only implied; the language used may be foreign, metaphorical or otherwise not standard. The index terms, however, must be such that all like terms are filed together and are normalized and cross-linked so that all rational approaches to the index will lead to the information sought. A list of such approved terms and their cross references is called an authority list or, more loosely, a thesaurus. Even in the simplest index it is advisable to have a list of the terms used in the index as a guide in the selection of index entries for new documents and as a guide in the selection of search terms. In large indexes it is mandatory that an authority list or thesaurus of index terms be maintained in order to avoid the scattering of entries due to the inadvertent use of synonyms. A thesaurus is also a valuable guide for selecting the cross references which should be searched.

Auto-Encoding and Keyword in Context (KWIC) Index

H. P. Luhn of IBM developed a system where a computer recognizes individual words and counts their frequency of occurrence in a text. Eliminating the very common words, such as articles, conjunctions,

prepositions, auxiliary verbs, and the like, the most commonly occurring words — the first 16 or so — could be used as index entries. This actually is a way of mechanically preparing a concordance. Although this is an indexing system, it was first used to select significant sentences for the preparation of abstracts, sometimes called auto-abstracts.

Auto-encoding is therefore a form of word indexing. Luhn recognized the limitations of word indexing and therefore undertook to standardize the vocabulary by combining words containing the same root and then combining the counts of words which are synonyms. These words are looked up in a thesaurus and a normalized form is substituted. Essentially, "normalizing" means selecting one form of a word so as to avoid scattering due to synonyms and inflected forms. In order to pinpoint more complex concepts, the computer would also analyze the text for word pairs, that is, cases where the statistically significant words followed each other in a sentence.

An inquiry into such an index could also be similarly constructed. The question would be in essay form and the search terms would be computer-developed from this essay, as were the index terms, and the two sets of terms would be matched.

Such a statistical analysis of words and word pairs, normalized by a thesaurus, is an experimental approach for the mechanized development of subject indexes.

One immediate practical benefit of this approach has been the development of the Key Word in Context (KWIC) Index. The process is applied to a title of an article or its abstract. All nonsignificant or "common" words are ignored and the remaining significant words, called keywords, are put in a fixed position within the title or sentence and arranged in alphabetic order:

KWIC INDEX

SIS AND RADIO- CHLORIDE	EXCHANGE OF TWO ALKYL CHLORIDES.	
ON RESPIRATORY GASEOUS	EXCHANGE OF TWO CHLOROPHYCED ALGAE (	VILLES-60-AAM
ION	EXCHANGE PROPERTIES OF KAOLINITE SLU	TAMEMA-60-IEP
ISOTOPIC	EXCHANGE REACTIONS. HALOGEN EXCHANG	HERBRH-60-IER
IONS OF COMPLEX IONS.	EXCHANGE REACTIONS OF PLATINUM(II) C	PEARRG-60-MSR
	EXCHANGE REACTIONS OF ALKALINE IONS	CHENEJ-60-ERA
ION	EXCHANGE RESINS AND THYROID METABOLI	BLANQP-60-IER
ION	EXCHANGE RESINS AS CATALYSTS.	KRESTR-60-IER
PREPARATION OF ION	EXCHANGE RESINS BY PEARL POLYCONDENS	LESEKF-60-PIE
OPTICAL OBSERVATION OF	EXCHANGE SPLITTING IN YTTERBIUM IRON	WICKKA-60-OOE
DEUTERIUM- HYDROGEN	EXCHANGE WHEN SOLID OLEFINS ARE BEIN	PONOAN-60-DHE
RAMAGNETIC RESONANCE OF	EXCHANGE-COUPLED CHROMIUM(+3) PAIRS	RIMAIL-60-PRE
SUBSTANCES IN AN ANION	EXCHANGER.	LINDLF-60-WSN
N A SCRAPING-BLADE HEAT	EXCHANGER.	ABRAMA-60-SST
DIDE- NUCLEOTIDE IN ION	EXCHANGER FRACTIONS.	MANNS -60-DDP
CRYSTALLIZATION IN IRON	EXCITATATION CURVES OF THE REACTIONS	JAOUMB-60-PDR
FOR THE PROBABILITY OF	EXCITATION BY ELECTRON IMPACT.	DORMFH-60-TLP
ERNAL CONVERSION OF THE	EXCITATION ENERGY. N- HETEROAROMATI	HOCHRM-60-LCM
TRAVIOLET RADIATION AND	EXCITATION OF OXYGEN LINES IN THE CH	NIKOGM-60-URE
PECTRA FOR FLUORESCENCE	EXCITATION OF PYRIDINE NUCLEOTIDE IN	OLSOJM-60-ASF
	EXCITATION OF THE AURORAL GREEN LINE	YOUNRA-60-EAG
E, GLUCOSE, AND MANNOSE	EXHIBITED.	EAGORG-60-EPA
ELECTRON MICROSCOPIC	EXHIBITION OF VERY LONG LINKS IN THE	HESSK -60-EME

Another arrangement is to put the keyword in a left-hand column and print out the whole title to the right, with or without the full bibliographic reference:

exchange	Isotopic exchange reactions. Halogen exchange in the system boron trichloride-phosphoryl chloride. J. Am. Chem. Soc., 82, 792-5 (1960)
exchange	Mechanism of substitution reactions of complex ions. Exchange reactions of platinum (II) complexes in various solvents. J. Am. Chem. Soc., 82, 787-92 (1960)
exchange	Ion exchange resins and thyroid metabolism. Quantitative determinations bearing on the plasma. Compt. Rend., 250, 218-19 (1960)
exchange	Ion exchange resins as catalysts. Ind. Chemist, 16, 3-8 (1960)

A KWIC index can be compiled very rapidly from machine-readable texts. It is, therefore, being used as a means of rapidly preparing and promptly disseminating bibliographies and announcement bulletins of new articles, books and other publications. The emphasis here, it should be noted, is on the dissemination of information and not on information retrieval. For this reason the KWIC index is sometimes referred to as a dissemination index rather than a retrieval index.

Recently, however, normalized keywords have been used as retrieval devices and tests are being made to check their effectiveness. The suspicion is growing that the machine-generated keywords, provided they are normalized and controlled, are just as effective in retrieving information as are subject headings, standard descriptors and classification schemes.

INDICATIVE AND INFORMATIVE INDEXES

Until recently, the purpose of indexing was merely to indicate where information about a subject may be found. The index did not tell the searcher what the information actually was. An index which just tells the general subject matter of a publication is called a descriptive or indicative index. An index which tries to give the searcher some idea of the contents of the publication is called an informative index. Some people have referred to the former as subject or document retrieval and the latter as data retrieval. In the former, it is the document as a whole, or a definite segment thereof, which is indexed. In the latter, it is the indexable item, the specific fact, which is indexed.

Most libraries have used indicative indexes and have confined themselves to subject retrieval. Serially published indexes, such as the abstract journals and indexes of research reports, which index in great detail with a high degree of specificity, try to provide informative indexes and, more recently, actually record exact data.

With an indicative index, it is sufficient, for example, simply to record that a certain document has information on the psychological effects of chemicals on schizophrenia. An index can, therefore, provide subject headings:

Psychological effects - Chemicals
Schizophrenia, Effect of Chemicals on

or, using coordinated indexing, enter the document under:

Schizophrenia
Chemicals
Psychology

If, however, an informative index is called for, it is necessary to be much more specific. What chemicals are involved? Which chemicals had an effect and which did not? Is this an actual experiment or is it a theoretical analysis? Are the chemicals used in conjunction or individually? And so on.

The nouns first have to be expanded. Instead of having Chemicals as a descriptor, one must replace it with the specific names. We shall code these as A, B, C, D; Psychological effects will be E and Schizophrenia F. In addition to expanding these nouns, we must show the relationships between them. Were the chemicals A, B, C, D used together or individually? In other words, is it A and B and C and D (a logical conjunctive) or is it A or B or C or D (a logical disjunctive)? Was there an effect or did the specific chemical have no effect? What is the subject and what the object? Or, differently, what is the causative agent and what is the resultant? There is not much chance for confusion here, but in many instances the difference would not be clear.

In other words, the syntactic relationships between terms must be expressed. In subject headings this can be shown by including the verbs

and prepositions necessary to show all relationships. For a complex subject, as in the example, this becomes too clumsy. It is more usual to supply a brief summary or informative note. A Chemical Abstracts type of index would read:

Chemical A
    Experimental tests showing effects on schizophrenia

Psychological effects
    Chemical A tests show effects on schizophrenia

Schizophrenia
    Chemical A tests show psychological effects

In a coordinate index, modulants (role indicators, interfixes) must be supplied to show the precise context and to show the interrelationship of the descriptors. A code for modulants can be easily developed by simply making a list of relationships and numbering them. For example:

and	101
or	102
source, induced by, produced by	103
intermediate material	104
final product, result, effect	105
agent, (solvent, catalyst, adsorbent)	106
increased by	107
decreased by	108
inhibited, blocked, arrested	109
compared with	110
test	111
inference, hypothesis	112

and so on.

These numbers can be used as the codes or symbols which are attached as modulants to the descriptors. In one system the coined word structerm has been applied to this combination of descriptor and role indicator. (Role indicators are also discussed in the section Coordinate Indexing.)

Again it must be emphasized that the use of modulants or role indicators is required only in certain instances. One can have extremely detailed indexing and get actual data retrieval with such descriptors as:

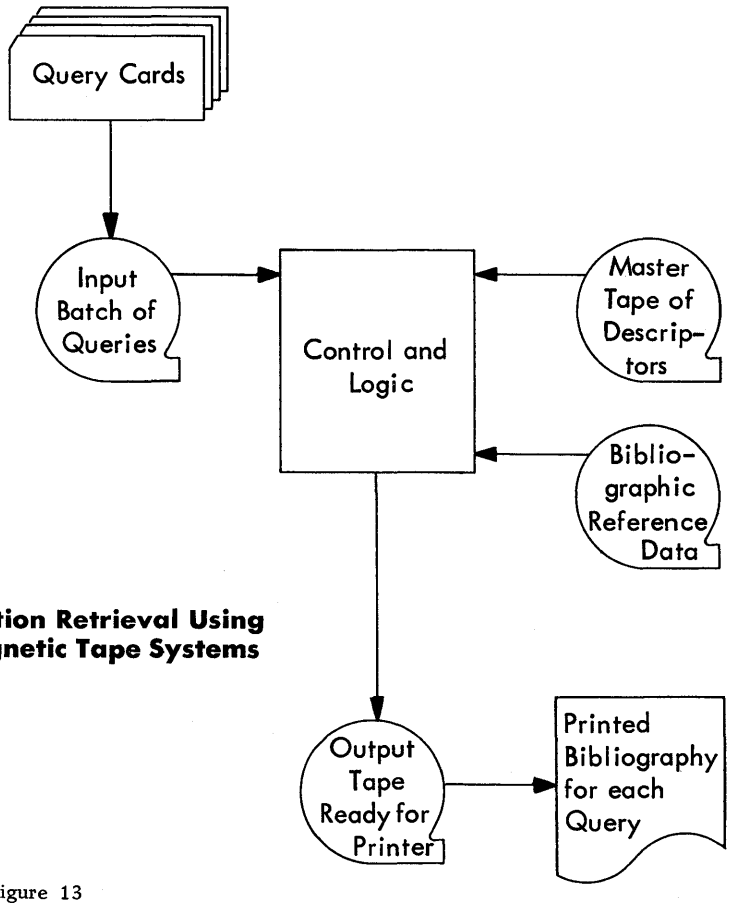
Aluminum
Tensile strength
800° F

These terms will retrieve a document that will give the tensile strength of aluminum at a specific temperature. In fact, the actual tensile

strength — that is, the specific data itself — can be incorporated with the document number on the descriptor 800° F. Thus indexes have been built which have provided both document (subject) retrieval and data retrieval. Referring to the arrangements on the punched cards or the magnetic tape, these index files are sometimes referred to as unformatted and formatted files respectively.

Indexes have almost always been systematically arranged either by class or alphabet. To find all the references, one simply looked up the appropriate headings and found the entries under each. To look up the entries with a machine requires either a random access device or the prefiling of cards, so that the individual decks may be read or collated separately.

The original data processing machines have, however, been serial machines. To work most efficiently they read the whole store and selected that which is needed. Such serial searching is inherently less efficient than lookup which provides more direct access. However, in many instances serial machines can perform information retrieval tasks more economically and faster than randon access machines.



Information Retrieval Using IBM Magnetic Tape Systems

Figure 13

Serial searching also provides additional benefits in that file maintenance is no problem, multiple entries do not have to be provided and input is greatly simplified. Searching also permits the infinite permutation of headings. The ability to relate various terms and to search for such relationships is materially facilitated in searching systems. This has

been especially useful, for example, in indexing the substructures of organic chemical compounds. Since it is impossible to anticipate all the possible combinations of chemical substructures for which one might wish to search — in other words, since there is a possibility of an almost infinite permutation of terms to be searched — a complete serial search is necessary.

With the current development of random access machines, however, especially units with large memories, the lookup principle is finding wider applicability. This is especially true for large files or files which will grow indefinitely. Also, current developments in using a thesaurus approach for cross references and for finding related subjects is making random access very attractive. Some systems use random access for current materials and store the older information on tape. Thus the material which is consulted frequently is more readily available while the less frequently used material is kept in more economical storage.

In setting up any index, the two systems, lookup and search, must be carefully compared and evaluated. Close attention must be paid to the size of the file, the length of search and the total costs in money, time and effort, as well as the inherent efficiencies of retrieving the desired information. Indexing systems should be developed first to solve the specific retrieval problem and without bias for a preconceived mechanical process. The adaptation of the system to a specific piece of equipment should then be accomplished.

APPLICATIONS

The emphasis throughout this discussion has been on the retrieval of information from existing documents. Furthermore, the examples chosen have been generally from science and technology. Information storage and retrieval, however, goes far beyond the handling of actual documents. The terms, Information Storage and Retrieval, really do not adequately describe the subject.

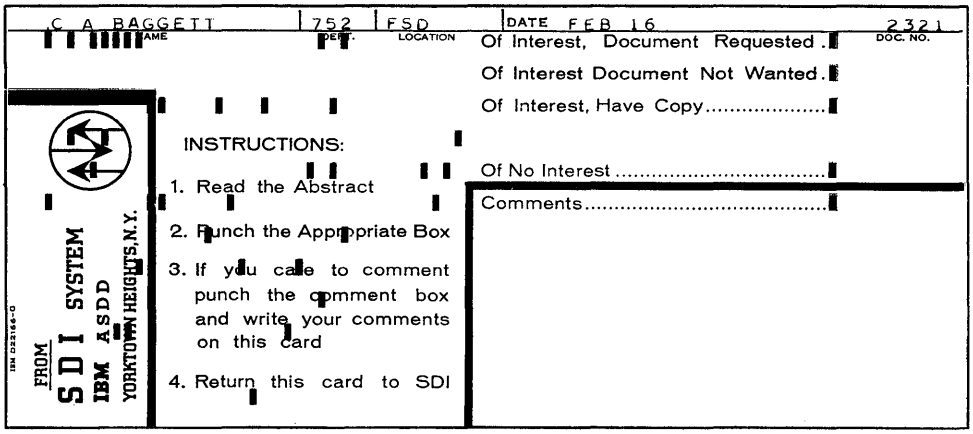
Storage and retrieval techniques are employed for all types of documents: engineering drawings, photographs, maps, licenses, insurance policies, correspondence, as well as books, reports, and other publications. The same techniques are also used for information that is not formally recorded in any document: personnel information, programmatic information and the like.

Similarly, information storage and retrieval must not be considered just as a passive activity responding to specific requests. It also has an active, dynamic function in disseminating information.

Information storage and retrieval is used, for example, to record information about developing programs. Progress and status of research and development activities, of construction, of business conditions, of stock levels, are all examples of programmatic information. The fast response required from such dynamic systems makes them proper candidates for mechanization.

There are, too, extensive informational needs about people. Records must be kept on skills, interests, pay rates, and other personnel matters. Health records, criminal records, driver's licenses, insurance policies, immigrant status and a host of other social records are currently being manipulated with punched cards and in computers. Programs have been developed for matching skill and interest indexes (profile registers) against document indexes in order to determine the individuals who should receive the new incoming information. This Selective Dissemination of Information (SDI) can be applied equally well in all fields and disciplines: science, technology, business. (See Figure 14.)

The principles developed for what are usually considered library applications can thus be applied to a host of problems that seemingly are remote from the library. In every instance, what is needed is a careful analysis of the desired end result in order to properly select the principles which should be applied.



C A BAGGETT | 752 | FSD | DATE FEB 16 | 2321
NAME | DEPT. | LOCATION | DOC. NO.

Of Interest, Document Requested
Of Interest Document Not Wanted
Of Interest, Have Copy
Of No Interest
Comments

INSTRUCTIONS:

1. Read the Abstract
2. Punch the Appropriate Box
3. If you care to comment punch the comment box and write your comments on this card
4. Return this card to SDI

FROM
SDI SYSTEM
IBM ASDD
YORKTOWN HEIGHTS, N.Y.

VHF and UHF Television Equipment #2321

William O. Swinyard June 1960
Proceedings of the IRE Vo. 48, No. 6, Pt. I

This paper covers a study of various types of VHF and UHF television receiving equipment made by TASO Panel 2 and reported October 3, 1958. Information and performance data are given for antennas, transmission lines and television receivers. RF amplifier and oscillator electron devices (tubes and semiconductors) used in television tuners for both VHF and UHF are discussed and tables showing relative performance data for devices of various types are included. Hard copies are limited. 15 pages

Figure 14

GLOSSARY

ABSTRACT, n. An epitome or summary of a document. An abstract may be locative, illative, indicative or informative. A locative abstract (used solely in a few legal libraries) specifies the place where the original document may be found. An illative abstract (used solely in a few legal libraries) specifies the general nature of the material in the document. An informative abstract includes and specifies all pertinent material in the original document. An indicative abstract points out what is in the original document, but usually does not include the material.

ACCESS, n. (1) A device or method whereby a document may be found. (2) Permission and opportunity to use a document.

ACCESSION, v. To register acquisitions.

ADDED ENTRY. A secondary entry — that is, any entry other than the main entry.

ADDRESS, n. (1) A label, name or number which designates a register, a location, or a device in a computer where information is stored. (2) That part of an instruction in a computer program which specifies the register, location or device upon which the operation is to be performed.

ALPHABETIC SUBJECT CATALOG. A catalog limited to subject entries and the necessary references, alphabetically arranged.

ALPHABETICO-CLASSED CATALOG. A catalog with entries under broad subjects alphabetically arranged and subdivided by topics in alphabetic order.

ALPHAMERIC, adj. (sometimes also Alphanumeric.) Expressed as either letters of the alphabet, numerals or special symbols.

ANALYTICAL SUBJECT ENTRY. A subject entry for part of a work, sometimes also called an Analytic.

ANNOTATED, adj. Supplied with annotations — that is, critical notes and commentaries.

APERTURE CARD. A punched card with an opening specifically prepared for the mounting of a frame or frames of microfilm.

AREA SEARCH. Examination of a large group of documents to segregate those documents pertaining to a general class, category, or topic. Screening.

ARRAY, n. (1) An ordinal arrangement of informational materials. (2) A set of mutually exclusive coordinate subclasses totally exhaustive of a class, derived by its division according to some one characteristic.

ASSOCIATION LINK. See Interfix.

ASYNDETIC, adj. Without cross references, said of a catalog.

AUTHOR ENTRY. Catalog entry under the name of the author, or under the heading which, according to the rules for author entries, corresponds to it.

AUTHOR NUMBER. See Book number.

AUTO-ABSTRACT, v. To select an assemblage of keywords from a document, commonly by an automatic or machine method, in order to form an abstract of the document.

AUTO-ENCODE, v. To select keywords from a document, by a machine method, in order to develop search patterns for information retrieval.

AUTO-INDEX, v. To select keywords from a document by a machine method in order to develop index entries.

AUXILIARY PUBLICATION. The process of making data available by means of specially ordered microfilm or photocopies. Auxiliary publication usually presupposes that the materials have not been published before, though it is sometimes applied to publication of Microcard copies of out-of-print books.

AUXILIARY SYNDESIS. The accessory apparatus — e.g., cross reference — which is used to supplement indexing sequence so as to reveal other relations.

BATCH PROCESSING. A technique by which items to be processed in a data processing machine must be collected into groups prior to their processing; contrasted to in-line processing.

BATTEN SYSTEM. A method of indexing invented by W. E. Batten, utilizing the coordination of single attributes to identify specific documents. Sometimes called the "peek-a-boo" system because of its method of comparing holes in cards by superimposing cards and checking the coincidence of holes.

BIBLIOGRAPHY, n. (1) An annotated catalog of documents. (2) An enumerative list of books. (3) A list of documents pertaining to a given subject or author. (4) The process of compiling catalogs or lists.

BIT, n. (1) Abbreviation of "binary digit." (2) A single character of a language employing exactly two distinct kinds of characters.

BLOCK INDEXING. A system of indexing wherein "blocks" of materials are collected, each block being small enough to permit easy manual search of the group contained therein.

BOOK NUMBER. A symbol, usually consisting of a combination of letters and figures, which serves to identify a given book among others bearing the same class number, and, at the same time, to place books bearing the same class number in the desired order on the shelves, by author, title, edition, and the like. When used to arrange books alphabetically by author, it is called author number or author notation.

BOUND, adj. (coordinate indexing) Joined in modification of the meaning of a commonly used term. For example, Free Energy is a bound term (unit concept) while Free and Energy may be free terms in the same coordinate indexing system.

BROWSABILITY, n. The ability of an indexing system to lend itself to unsystematic or random searches. This ability is of interest or use to the searcher even though it may not produce a logical answer to the search question.

BROWSE, v. To investigate, without design, the contents of a collection of books or documents.

BRUSSELS CLASSIFICATION. The Universal Decimal Classification.

BUCKET, n. A section in the memory of a computer.

CALENDAR, n. A chronologically arranged sequence of documents pertaining to a single author, subject, series or class.

CALL NUMBER. The class number and the book number by which the location of the book on the shelf is indicated.

CARD CATALOG. A catalog made up of cards, each usually bearing a single entry. The card catalog is to be distinguished from the printed catalog, in book form, and the sheaf catalog, which consists of sheets brought together in portfolios.

CATALOG, n. A register or compilation of items arranged methodically, usually with sufficient description to afford access.

CATALOG, v. To register or compile a list of documents with sufficient description to afford access.

CATALOGER, n. One who catalogs, as books or documents.

CATCHWORD (SCHLAGWORT) INDEX. One which uses a significant word from a title or text to index an item.

CATEGORY, n. (1) A comprehensive class or description of things. (2) A logical grouping of associated documents. (3) A class or division formed for purposes of a given classification. In faceted classification special distinctions are made between categories, classes, facets and phases.

CHAIN INDEX. An alphabetic index wherein a heading is provided for each term or link for all the terms used in a subject heading or classification. See also Relative index and Correlative index.

CHECK LIST, n. An enumeration of documentary holdings with a minimum of organization and bibliographic information.

CHECKOUT ROUTINE. (1) A procedure used in machine documentation systems to determine the correctness of answers, involving the use of sample inquiries, the answers to which are known. (2) The necessary procedures demanded before removing a document from a collection.

CLASS, n. (1) A group having the same or similar characteristics. (2) A major subdivision of a category.

CLASS NUMBER. A symbol applied to a book, etc., indicating the class to which it belongs in the classification system used by the library.

CLASSED CATALOG. A catalog arranged by subject according to a systematic scheme of classification. Also called "class catalog," "classified subject catalog," and "systematic catalog."

CLASSIFICATION, n. A distribution into groups. A systematic division of a group of related subjects. A schedule for the arrangement or organization of documents.

CLASSIFICATIONIST, n. One who makes classification schedules. A theorist who organizes and divides documents according to a specific criterion.

CLASSIFIED INDEX. An index characterized by subdivisions of hierarchic structure. An index using or displaying genus-species (class-subclass) relationships. Cf. Classed Catalog.

CODE, n. (1) A communication system for information. (2) A system of symbols used in transmitting or storing information. (3) System of arbitrary signs and symbols used to represent words or concepts, as distinguished from a cipher wherein arbitrary signs and symbols are used to represent single letters or syllables. (4) A systematic body of laws, regulations or rules.

COLLATE, v. (1) To compare or examine critically, particularly to verify the presence or absence of specific items in a text, for example, printer's errors, missing pages, handwritten annotations. (2) To assemble the pages of a document in correct order — hence, also, to interleave. (3) To merge and combine two or more similarly ordered sets of items to produce an ordered set.

COLON, n. (1) A device used in the U. D. C. to link related class terms. (2) A device used in the Colon Classification to separate successive foci. Later, in the Colon Classification, a device to introduce the energy facet.

COLON CLASSIFICATION. A faceted classification scheme developed by S. R. Ranganathan.

COMPENDIUM, n. An abbreviated summary of the essentials of a subject — specifically, a book containing such treatment.

CONCEPT COORDINATION. A system of multidimensional indexing with single concepts to define a document uniquely. Cf. Coordinate Indexing, Uniterm Indexing, Zatocoding System.

CONCORDANCE, n. An alphabetic list of words and phrases appearing in a document, with indications of the context of such words and phrases in the text.

CONJUNCTIVE, adj. Pertaining to the joining or coupling of two documents, words, phrases, or elements of information in order to express a unity. Being neither disjunctive nor collateral.

COORDINATE INDEXING. An indexing scheme whereby the interrelations of terms are shown by coupling individual words. Cf. Manipulative Index, Uniterm Indexing, Zatocoding System.

CORPORATE NAME. The name of a corporate body as distinguished from the name of a person.

CORRELATION, n. A systematic or reciprocal connection — sometimes, the establishment of a mutual or reciprocal relation of or between.

CORRELATIVE INDEX. An index enabling selection of documents or of references to them by correlation of words, numbers, or other symbols which are usually unrelated by hierarchic organization.

CROSS REFERENCE. A reference or direction made from one term or one part of an index to another related term or part.

DECIMAL CLASSIFICATION. See Dewey Decimal Classification.

DECK, n. A collection of cards, commonly a complete set of cards, which have been punched for a definite service. (In Britain, the more common term is pack.)

DECKLET, n. A set of cards forming a single record.

DESCRIPTOR, n. (1) An elementary term. (2) A simple word or phrase used as a subject.

DEWEY DECIMAL CLASSIFICATION (DC). Classification system developed by Melvil Dewey and used very extensively for the shelf arrangement of books.

DICTIONARY, n. (1) Words arranged alphabetically and usually defined. (2) A lexicon in alphabetic order.

DICTIONARY CATALOG. A catalog in which all entries are interfiled to form a single alphabet, as in a dictionary.

DOCUMENT, n. An instrument having recorded information regardless of its physical form or characteristics.

DOCUMENT CARD. A Unit card, which see. A card carrying all the bibliographic and index information for an item. Used in Zatocoding and other edge-notched card systems as well as in serially searched files.

DOCUMENTATION, n. (1) The science of collecting, storing and organizing recorded informational materials or documents for optimum access. (2) "Includes the activities which constitute special librarianship plus the prior activities of preparing and reproducing materials and the subsequent activity or distribution." (3) "Selection, classification, and dissemination of information." (4) "The science of ordered presentation and preservation of the records of knowledge serving to render their contents available for rapid reference and correlation." (5) "The procedure by which the accumulated store of learning is made available for the further advancement of knowledge." (6) "The art of facilitating the use of recorded, specialized knowledge through its presentation, reproduction, publication, dissemination, collection, storage, subject analysis, organization, and retrieval." (7) "Collection and conservation, classification and selection, dissemination and utilization of all information."

DUPLICATE ENTRY. Entry of the same subject matter under two distinct aspects of it.

EDITION, n. The whole number of copies of a publication printed at any time or times from one setting up of type. An impression, issue or printing is the whole number of copies printed at one time.

ENCODE, v. (1) To put in symbolic form. (2) To transform a document, message or abstract by means of a specific notation.

ENTROPY, n. The unavailable information in a group of documents. The degree of disorganization in an informational assemblage.

ENTRY, n. A record of a document in a catalog, list or index.

ENUMERATIVE CLASSIFICATION. A classification based on a list of the individual subjects to be included.

EPITOME, n. A concise summary; a brief statement of the contents of a work.

FACET, n. An aspect or orientation of a topic.

FACETED CLASSIFICATION. Classification schemes whose terms are grouped by conceptual categories and ordered so as to display their generic relations. These categories or "facets" are standard unit-

schedules and the terms, or rather the notation for the terms from these various unit-schedules, are combined at will in accordance with a prescribed order of permutation or combination.

FACSIMILE, n. A precise reproduction of an original document; an exact copy.

FALSE-DROP, n. Citation that does not pertain to the subject sought. An alien, usually in a manipulative or coordinate index.

FEEDBACK, n. Partial reversion of the effects of a given process to its source. Control of a system by the output of the system — that is, a self-correcting or self-compensating control.

FICHE, n. A card (European usage).

FIELD, n. A fixed column or group of columns in a punched card allocated for punching specific information. The total area of a punched card available for information storage.

FILE, n. An organized collection of information directed toward some purpose.

FILMOREX SYSTEM. A system for the electronic selection of microfilm cards devised by Jacques Samain. Each card has a micro-reproduction of the document or abstract and a field of twenty 5-digit code numbers giving the bibliographic reference and the subjects treated.

FREE, adj. (coordinate indexing) Alone, not bound or joined to a separate modifier. (See Bound.)

GAP, n. A hiatus in a collection, commonly of serials or regularly issued proceedings.

GENERAL REFERENCE. A blanket reference in an index or catalog to the kind of heading under which one may expect to find entries for materials on certain subjects or entries for particular kinds of names. Also called "general cross reference" and "information entry."

GENERIC, adj. Pertaining to a genus or class of related things.

GENUS, n. A class of similars divisible into two or more subordinate classes or species.

GLOSSARY, n. (1) An explanation of the meanings of terms peculiar to a subject field. (2) A collection of equivalent synonyms in two or more languages.

HARD COPY, n. A human-readable copy produced from information that has been transcribed to a form not easily readable by human beings.

HAYSTAQ. Name of an information searching procedure with electronic computers used by the U. S. Patent Office.

HEADING, n. The word, name or phrase at the beginning of an entry to indicate some special aspect of the document (authorship, subject content, series, title, etc.).

HIERARCHIC, adj. (1) Arranged in serial rank rather than ordinal position. (2) Pertaining to a generic classification or organization of materials.

HIT, n. Term used in mechanized retrieval systems to represent an apparent answer found by the machine.

HOLOGRAPH, n. A manuscript or document wholly in the author's own handwriting.

IMPRINT, n. (1) The place of publication, the name of the publisher and the date. (2) The title, author and other information stamped on the spine of a book.

INDEX, n. That which specifies, indicates or designates the information, contents or topics of a document or a group of documents. Also, a list of the names or subjects referring to a document or group of documents.

INDEX, v. To prepare an organized or systematic list which specifies, indicates or designates the information, contents or topics in a document or group of documents.

IN-LINE PROCESSING. A technique by which an item may be fully processed, with random access to all of the entries which that item may affect. The processing of data without sorting or any prior treatment other than storage.

INPUT, n. That which is put in — that is, the information transferred from external storage to the internal storage of the machine.

INTERCALATE, v. To file or insert, as in a card catalog.

INTERFIX, n. A device to signal relationships between concepts. Thus for a series of compounds, A, B, C.... insertion of the interfixes 1 and 2 (for example, A_1, B_1, B_2, C_2 ....) signals that the compounds with the same numerical interfix are in one mixture and those with a different one are in a different mixture. Cf. Role Indicator, Modulant.

ITEM, n. In an index, the reference to the document. Cf. Term.

KEYTERM, n. See Keyword.

KEYWORD, n. Grammatical element which conveys the significant meaning in a document. Word indicating a subject discussed in document.

KEYWORD IN CONTEXT INDEX. A listing, usually of titles or significant sentences from an abstract, with the keywords put in a fixed position within the title or sentence and arranged in alphabetic order in a column.

KWIC INDEX. An abbreviation for Keyword in Context Index.

LATTICE, n. The network of interrelationships between specific subjects.

LEXICON, n. An ordered vocabulary with definitions. When the alphabet is used to order the vocabulary, the lexicon is a dictionary.

LIBRARY OF CONGRESS CLASSIFICATION (LC). Classification scheme developed by the Library of Congress to arrange its collection.

LITERATURE SEARCH. A systematic and exhaustive search for published material bearing on a specific problem or subject, with the preparation of abstracts for the use of the researcher; an intermediate stage between reference work and research, and to be differentiated from both.

LOG, n. A registry of items, e.g., an accession list.

MACHINE-LANGUAGE CODING. (1) Linguistic or numerical patterns susceptible of being handled by data processing equipment. (2) A special type of notation used for a specific data processing machine. (3) Coding in the form in which instructions are executed by the computer. Contrasted to relative, symbolic, and other non-machine-language coding.

MAIN ENTRY. A full catalog entry, usually the author entry, giving all the information necessary to complete identification of a work. In a card catalog this entry bears also the tracing of all the other headings under which the work in question is entered in the catalog.

MANIPULATIVE INDEX. An index in which manipulations other than turning pages, reading entries, following cross references, and locating documents are necessary. Mechanized indexes using punched cards, and the various coordinate indexing systems are examples.

MARK-SENSE, v. To indicate a punch position by means of an electrically conductive pencil mark in such a way that a suitably designed machine can make the punch automatically.

MERGE, v. Combine two files, already in sequence, into a single file.

MICROCARD, n. (1) An opaque photographic reproduction generally not readable without optical aid. (2) Both opaque and transparent photographic reproduction. (3) Trade-mark of the Microcard Corporation. Cf. Microfiche.

MICROCOPY, n. A facsimile of substantially reduced size. A microtext.

MICROFICHE, n. A set of microphotographs on sheet microfilm, usually about 3 x 5 inches or 9 x 12 cm.

MICRO-OPAQUE, n. An opaque microcopy, such as microcards.

MICROPHOTOGRAPH, n. A reduced-size photographic documentary reproduction generally too small to be read with the unaided eye.

MICROPHOTOGRAPHY, n. The process of making a very small photograph of a much larger original.

MICROPRINT, n. Printing, reproductions of printing or other documents of reduced size on opaque paper.

MICROTEXT, n. A documentary facsimile of substantially reduced size. A microcopy.

MICROTRANSPARENCY, n. A transparent microcopy.

MINICARD. An Eastman Kodak Corporation system for information storage and retrieval. Documents and digital dot codes are photographed onto film chips 5/8 in. x 1 1/4 in. These chips are held on rods or "sticks" for transport, storing and feeding them. Chips are searched by sorting and copies prepared from document images selected.

MINITEXT EDITION. Microprint version of a document whose text layout has been rearranged to fit a given size page.

MODULANT, n. An Interfix; a standardized suffix added to the root of a word (Ruly English) to bring out the different aspects of a word's basic meaning (U. S. Patent Office).

MULTI-ASPECT INDEX. See Coordinate Indexing.

NOISE, n. An undesirable signal which disturbs the desired signal in a communication network. See False-drop.

NOTATION, n. An arbitrary device to indicate the contents or location of a document.

OFFPRINT, n. A separate, an excerpt, as a magazine article, separately printed.

OPEN-ENDED, adj. Being possessed of the quality by which the addition of new terms, subject headings, or classifications does not disturb the pre-existing system.

OPERATION CODE. That part of an instruction in a computer program designating the processing step to be performed.

OUTPUT, n. The product of a process — that is, the information transferred from the internal storage of a computer to output devices for external storage.

PACK, n. A collection of cards. (See also Deck, which is the more common term in the U. S.) Usually refers to a complete set of cards which have been punched for a definite service.

PAMPHLET, n. A short work commonly bound as a single fascicle, and published as a separate issue. Unlike a reprint or separate, a pamphlet is not a part of a larger work.

PARALLEL, adj. Pertaining to the simultaneous handling of all the elements of a group. Cf. Serial.

PEEK-A-BOO SYSTEM. See Batten System. Includes also commercial variations such as Cordonnier, Taylor, Matrex systems.

PERMUTATION INDEXING. See Rotational Indexing.

POST, v.t. (1) To transfer an indicial notation from a parent or main entry to individual analytic entries — for example, to type the proper catalog entry and number at the top of a group of catalog cards. (2) (coordinate indexing) To put the accession number of a document under each entry representing a coordination term.

POST COMBINATION INDEXING. See Coordinate Indexing.

PROGRAM, n. An outline giving the schedule of actions to be followed or the order and arrangement of such a schedule. A series of instructions expressed in symbols which a machine system can accept and understand.

PUNCHED CARD, n. A card of lightweight cardboard on which information is represented by holes punched in specific positions.

QUALIFIED HEADING. A heading followed by a qualifying term which is usually enclosed in parentheses, e.g., Composition (Art), Composition (Law).

RANDOM ACCESS STORAGE, n. A storage technique in which the time required to obtain information is independent of the location of the information — that is, items do not have to be processed in sequence.

RANK, n. A measure of the relative position in a series, group, classification or array.

RAPID SELECTOR. A machine for document storage and retrieval. Documents are photographed onto 35mm microfilm and alongside are placed digital dot codes indexing each frame. In searching, a reel of film is run past an optical scanner which reads the optical dot pattern. Documents selected are copied automatically from the reel of film.

REDUNDANCY, n. Use of more words than needed to convey the thought. An excess of rules and syntax whereby it becomes increasingly likely that mistakes in reception will be avoided.

"REFER FROM" REFERENCE. An indication, in a list of subject headings, of the headings from which references should be made to the given heading; it is the reverse of the indication of a "see" or "see also" reference.

REFERENCE, n. (1) A direction from one heading to another. (2) An indication referring to a document or passage.

RELATIVE INDEX. An alphabetic index to a classification scheme in which all relationships and aspects of the subject are brought together under each index entry.

RETRIEVAL, n. The act of finding again, recovery, retrospective searching and securing of documents. The act of going to a specific location or area and returning therefrom with an object or document.

ROLE INDICATOR. See Interfix.

ROTATIONAL INDEXING. Correlative indexing (which see) wherein each term is "rotated" so as to file in the first position.

RULY ENGLISH. English in which every word has one and only one conceptual meaning and each concept has only a single word to describe it. Terms proposed by S. Newman of the U. S. Patent Office to develop certain index codes.

SCAN, v. To examine every reference or every entry in a file routinely as part of a retrieval scheme.

SCAN-COLUMN INDEX. A coordinate book-form index developed by J. O'Connor which provides for manual serial searching of terms arranged in columns.

SCOPE NOTE. A statement giving the range of meaning and scope of a subject heading or descriptor and usually referring to related or overlapping headings.

"SEE ALSO" REFERENCE. A reference to a less comprehensive or otherwise related term.

"SEE" REFERENCE. A reference from a term or name under which no documents are entered to that used in place of it.

SELF-ORGANIZING, adj. Capable of spontaneous classification.

SEMANTIC CODE. A linguistic system developed for use on machines designed to detect logically defined combinations; a symbol representing the concept of a word. (Used by J. W. Perry.)

SEMANTIC FACTORS. Generalized concepts used to construct Semantic Code. (Used by J. W. Perry.)

SEPARATE, n. A reprint or special copy of an article, chapter or other part of a larger publication. Distinguished from Pamphlet (which see) in having been issued originally in a larger publication.

SERIAL, adj. The handling of data in a sequential fashion. Cf. Parallel.

SERIAL, n. A publication issued in successive parts and implying perpetual continuation.

SHELFLIST, n. A record of the books in the library arranged in the order in which they stand on the shelf — that is, in the order of their class and book numbers.

SPECIFIC ENTRY. Entry of a document under a heading which expresses its special subject or topic as distinguished from the class or broad subject which includes that special subject or topic.

SPLIT CATALOG. A library catalog in which the different varieties of entry — e.g., subject, author, title — are filed in separate alphabets.

STORAGE, n. A source from which documents or information of specified descriptions may be supplied. A receptacle for information.

STRUCTERM, n. A term or descriptor having an appended role code indicating context in which term is used. (Used by F. R. Whaley.)

SUBJECT AUTHORITY CARD. A card which, in addition to citing the authorities consulted in determining the choice of a given heading, also indicates the references made to and from related headings and synonymous terms.

SUBJECT CATALOG. A catalog consisting of subject entries only.

SUBJECT HEADING. A word or group of words indicating a subject under which all material dealing with the same theme is entered in an index, catalog or bibliography, or arranged in a file.

SUPPLIED TITLE. The title composed by the cataloger to indicate the nature and scope of the monographic work under study.

SYMBOL, n. A substitute or representation of characteristics, relationships, or transformations of ideas or things.

SYNDETIC, n. Having entries connected by cross references. A coordination of two or more related documents.

SYNOPSIS, n. An essential summary of actions. In fiction, the argument of a story.

TABLEDEX INDEX. A coordinate book-form index developed by R. S. Ledley. Terms are arranged in tables with document numbers and associated term numbers in ascending number sequence.

TAPE, n. (data processing) (1) A plastic strip coated or impregnated with magnetic or optically sensitive substances, used for data input, memory or output. (2) A paper or plastic strip with punches or other arbitrary signs representing alphabetic or numerical data and operations.

TAXONOMY, n. The science of classification. Also, the study of the names and naming of items in generic assemblies.

TELEGRAPHIC ABSTRACTS. A special abbreviated style of abstract, commonly considered suitable for machine input. (Used by J. W. Perry.)

TELEREFERENCE, n. A method for consulting catalogs from a remote location, consisting of a closed-circuit television system for viewing the catalog, a relay for finding the part of the catalog to be examined, and mechanical handling equipment for moving the catalog cards or pages about.

TERM, n. In an index the subject heading or descriptor. Cf. Item.

THESAURUS, n. A lexicon, more especially where words are grouped by ideas; a grouping or classification of synonyms or near synonyms; a set of equivalence classes of terminology.

TITLE ENTRY. The record of a work in a catalog or bibliography under the title.

TRACING, n. In a card catalog, the record on the main entry card of all the additional headings under which the work is represented in the catalog. Also, the record on the main entry card or on an authority card of all the related references made. In coordinate indexing, a list of the descriptors, Uniterms, etc., applied to a specific document.

UNION CATALOG. An orderly compilation of the holdings of two or more libraries, presumptive of cooperation between the libraries.

UNIT CARD. A basic catalog card, in the form of a main entry, which, when duplicated, may be used as a unit for all other entries for that work in the catalog by the addition of an appropriate heading.

UNITERM INDEXING. A form of index display developed by Mortimer Taube which utilizes single descriptors, called Uniterms, to define a document and which facilitates the manual coordination of these descriptors. Cf. Descriptor.

UNIVERSAL DECIMAL CLASSIFICATION (UDC). An expansion of Dewey Decimal Classification started by P. Otlet in Brussels, sometimes referred to as the Brussels system.

WEED, v. To discard currently undesirable or needless materials from a file.

XEROGRAPHY, n. A dry copying process involving the photoelectric discharge of an electrostatically-charged selenium plate. The charge is "developed" by cascading a thermoplastic "toner" over the plate. The toner adheres to the image areas, the remaining electrostatic charge is discharged and the toner is transferred to paper or an offset printing master and then fused by heat.

ZATOCODING SYSTEM. A system of coordinate indexing developed by Calvin Mooers, using random superimposed coding on edge-notched cards.

BIBLIOGRAPHY

American Library Association. A. L. A. rules for filing catalog cards. Corr. ed., Chicago, A. L. A., 1943.

Andrews, Don D., and Newman, Simon M. Storage and retrieval of contents of technical literature; nonchemical information. Preliminary report May 15, 1956. Washington, U. S. Patent Office. Research and Development Report No. 1. Also: First supplementary report, June 1957. Research and Development Report No. 4. Second supplementary report, November 17, 1958. Research and Development Report No. 12.

Armed Services Technical Information Agency. Conference on multiple-aspect searching for information retrieval in Washington, D. C. February 12th and 13th, 1957. ASTIA report AD-147491.

Bar-Hillel, Yehoshua. A logician's reaction to recent theorizing on information search systems. American Documentation v. 8, 1957, p. 103-116.

Bernier, Charles L. Correlative indexes. American Documentation v. 7, 1956, p. 283-288; v. 8, 1957, p. 47-50, 211-220, 306-313.

Casey, Robert S., Perry, James W., Kent, Allen, and Berry, Madeline. Punched cards: Their application to science and industry, 2nd ed. N. Y., Reinhold, 1958.

Department of Health, Education and Welfare, Social Security Administration, Bureau of Old Age and Survivors Insurance, Division of Accounting Operations. Preliminary report of distribution of surnames in the Social Security account number file. 1957.

Dewey, Melvil. Decimal classification and relative index, 14th ed., 1946; 15th ed., 1951; 16th ed., 1958.

Haykin, David Judson. Subject headings, a practical guide. Washington, U. S. Government Printing Office, 1951.

Joyce, T., and Needham, R. M. The thesaurus approach to information retrieval. American Documentation, v. 9, 1958, p. 192-197.

Luhn, H. P. Auto-encoding of documents for information retrieval systems. Yorktown Heights, N. Y., IBM, 1958.

Luhn, H. P. The automatic derivation of information retrieval encodements from machine-readable texts. Yorktown Heights, N. Y., IBM, 1959.

Luhn, H. P. The IBM universal card scanner for punched card information searching systems. Yorktown Heights, N. Y., IBM, 1958.

Luhn, H. P. Keyword-in-context index for technical literature (KWIC) index. Yorktown Heights, N. Y., IBM, 1959. ASDD report RC-127.

Luhn, H. P. Row-by-row scanning systems for IBM punched cards as applied to information retrieval problems. Yorktown Heights, N. Y., IBM, 1959.

Luhn, H. P. Selective dissemination of new scientific information with the aid of electronic processing equipment. Yorktown Heights, N. Y., IBM, 1959.

Luhn, H. P. A statistical approach to mechanized literature searching. Poughkeepsie, N. Y., IBM, 1957. Research paper RC-3.

Mann, Margaret. Introduction to cataloging and the classification of books. 2nd ed., Chicago, A. L. A., 1943.

Metcalfe, John. Information indexing and subject cataloging. Alphabetical: Classified: Coordinate: Mechanical. N. Y., Scarecrow Press, 1957.

Metcalfe, John. Subject classifying and indexing of libraries and literature. N. Y., Scarecrow Press, 1959.

Modern Coding Methods. N. Y., IBM, n.d.

Mooers, C. N. Zatocoding and developments in information retrieval. ASLIB Proceedings v. 8, 1956, p. 3-22.

Murphy, R. W. The IBM 9900 Special Index Analyzer. Poughkeepsie, N. Y., IBM, 1958.

Nolan, J. J. Principles of information storage and retrieval using a large scale random access memory. San Jose, Calif., IBM, 1958.

Odell, Margaret K., and Strong, Earl L. Records management and filing operations. N. Y., McGraw-Hill, 1947.

Perry, J. W., and Kent, Allen. Tools for machine literature searching; semantic code dictionary, equipment, procedures. N. Y., Interscience, 1958.

Richmond, Phyllis A. Hierarchical definition. American Documentation. v. 11, 1960, p. 91-96.

Savage, T. R. The preparation of automatic abstracts on the IBM 704 data processing system. Yorktown Heights, N. Y., IBM, 1958.

Taube, Mortimer, and Wooster, Harold. Information storage and retrieval; theory, systems and devices. N. Y., Columbia University Press, 1958.

U. S. Library of Congress. Subject Cataloging Division. Subject headings used in the dictionary catalogs of the Library of Congress. 6th ed., edited by Marguerite V. Quattlebaum. Washington, 1957.

Wadington, John P. Unit concept coordinate indexing. American Documentation, v. 9, 1958, p. 107-113.

Wagner, Frank S., Jr. A dictionary of document terms. American Documentation, v. 11, 1960, p. 102-119.

Warheit, I. A. Evaluation of library techniques for the control of research materials. American Documentation, v. 7, 1956, p. 267-275.

Weeks, Bertha M. How to file and index. Rev. ed. N. Y., Ronald Press, 1956.

C20-8062

Printed in U.S.A. C20-8062

IBM

International Business Machines Corporation
Data Processing Division
112 East Post Road, White Plains, New York