State of the Art of Published Indexes
JOHN MARKUS
American Documentation (pre-1986); Jan 1962; 13, 1; ABI/INFORM Global
pg. 15

Despite the glamour of mechanized searching for needed information, there is still a real need for published indexes. These indexes are in effect miniature, portable information centers operating on a do-it-yourself basis. Published indexes have the important advantage of permitting browsing, which is highly essential when the searcher does not know exactly what he is looking for. They can also give needed information almost immediately because there is no waiting for search time on a computer.

Once a book-form index is acquired, its use costs nothing except time. The greatly increased output of published material in the past two decades has created new index-publishing problems. In this state-of-the-art survey, the emphasis is put on the new publishing techniques, new equipment, and innovations in index format that have been developed in attempts to solve these problems. Representative examples of each type of indexing venture are cited. The many other equally meritorious examples of new published indexes are omitted only because of space limitations.

State of the Art of Published Indexes

Received 16 November 1961

JOHN MARKUS

Technical Director
Technical Information Research Staff
McGraw-Hill Book Company
New York, New York

• Magnitude of the Index-publishing Problem

Today it is impossible for the average scientist or engineer to keep up with the published information in his specialized field. He must therefore fall back on some means of finding what he needs when he needs it. At the present time, published indexes are the best available means for achieving this goal for the average scientist who has no access to operational documentation centers.

Each year now there are about 1.5 million documents that should be indexed. How large would an annual index be for 1.5 million documents? Assuming a conservative 10 lines per document for subject and author entries and 300 lines per printed page, this dream index would require 50,000 pages. It would fit into 50 encyclopedia-size volumes, divided alphabetically as for encyclopedias, and would fill a 7-foot bookshelf each year. This is the picture to keep in mind as the various index-publishing problems and solutions are considered one by one.

As indexes grow in size, users complain about the increasing number of pages they have to flip through and the increasing time lags in publication of the indexes. Chopping up indexes into smaller specialized units is another alternative, but scientific literature does not always drop into discrete categories. Because of unavoidable overlap of fields, an increase in the number of specialized indexes would force users to go through many volumes to make an inter-disciplinary search. Increasing the number of indexes also increases the cost of piecemeal screening and sorting of documents into the assigned categories for specialized indexes.

• Index Cost Problems

Money is at the root of most index-publishing problems. Individuals rarely buy published indexes, partly because of their high and continuing cost and partly because they usually have to go to a library anyway to get the documents cited in an index. The market for index volumes is thus small because it consists largely of libraries. As one example, *Engineering Index* has a press run of only about 1,500 copies for its annual volume which now sells for $75. In contrast, most technical book manuscripts must show a potential market of well over 5,000 copies before they become attractive to a large technical book publisher.

As another example of the limited market for published indexes, certain McGraw-Hill publications make their annual index available separately to readers asking for it. Even when entirely free, with no index published in the magazine, there are rarely more than 2,000 requests.

Because of this economic situation, the current trend is toward Government or other subsidized support of new index projects. Commercial publishers do not like this, but their hands are tied because they have a responsibility of producing a profit for their stockholders.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

• Depth-of-indexing Problems

The optimum number of subject entries needed per document is a much-talked-of question in connection with indexing systems. With published indexes, the number of pages available generally governs the depth, because pages are directly related to cost. The annual index for a typical McGraw-Hill publication has an average of less than five subject and author index entries per article, because space is at a premium in magazines. The more generous Atomic Energy Commission budget for *Nuclear Science Abstracts* permits an average of about eight entries per document.

Where economics is not a dominating factor, depth of indexing should be determined by the needs of the intended users. The publishers of *Chemical Abstracts*, for example, feel that an average of 12 entries is required to meet minimum user needs in this field. Over-indexing can be bad because it leads the user to disappointingly trivial material in an article. There is thus no simple answer to the depth-of-indexing problem.

It is interesting to note that deep indexing costs little more than shallow indexing. The major portion of an indexer's time is used in getting to know the document. Once this is done, it takes little more time to mark or write down additional indexing terms.

• Index Quality Problems

Despite investigations by Cleverdon in England (1, 2) and by many others, there is today no generally accepted method of comparing the effectiveness of different types of indexes. The needs of index users vary so greatly that even the most carefully planned tests of retrieval efficiency can be challenged. An index which is perfect for one user can be almost worthless for another who has different needs.

Starting with indexers, some of the variables that complicate the problem of comparing indexing systems are the viewpoint and ability of the indexer, the depth of indexing used, and the number of cross-references used. At the user end of the problem, the variables include the viewpoint and searching ability of the user, the time available for making a search, the degree of motivation for searching, the readability of the published index, and the ease of browsing.

There are more and more indications that most indexing systems can be made to produce essentially the same degree of effectiveness. The real test is how well a given system serves its intended users.

When indexers have to use their own product for answering queries, they keep up to date with current requirements of index users and expose themselves to the inadequacies of their own indexing. This feedback increases the quality of any indexing system. With *Nuclear Science Abstracts,* feedback is achieved by rotating indexers and searchers at regular intervals.

An ideal index would guide the searcher to every relevant item of information in every document. Published indexes fall far short of this ideal. One reason is the high cost of using the number of entries required to give such perfection in indexing. Another is the lack of a crystal ball that tells what somebody will want to look up in the future. Finally, an ideal index serving all future needs would be greatly over-indexed for current use. As a result, most index publishers try to achieve a practical compromise that will please most of the users most of the time.

• Speed-of-publishing Problems

One common complaint about published indexes is that they lag too far behind material that is indexed. The amount of this lag is dependent on two factors: (1) The frequency of publication; (2) the time required to produce and publish the index after cutoff of incoming material.

With an annual index, the frequency lag will be a full year or more for some of the material. The only way to reduce this inherent lag is to increase the frequency of publication. This boosts publishing costs considerably, even when advantage is taken of mechanized techniques for cumulating the indexes.

With conventional printing techniques, composition cannot start until after the last entry has been received and alphabetized. The newly developed index-publishing techniques described in the following sections shorten the production lag by composing concurrently with indexing, sorting at high speed on machines, then producing negatives on card cameras operating at many times the speed of conventional typesetting machines.

• Keypunched Cards Speed Index Production

Keypunching of punched cards or paper tapes can minimize the repetitive manual operations involved in converting underlined, encircled, or handwritten indexing terms and accompanying bibliographic data to the final published title, author, subject, and other indexes. Each item of essential index data is keypunched only once, in a prescribed sequence. From this one master punched record, programmed automatic keypunching machines can reproduce this basic data to produce entries for the various types of indexes desired. The entries are alphabetized or otherwise sorted by machines, then converted to repro copy (final copy for reproduction) by a printer. The chief drawback of this technique is the fact that the printer output can at the present time be only all-caps, without proportional spacing of characters. This is hard to read and wasteful of space.

• Card Cameras Speed Index Production

Many indexes are today produced by typing reproduction copy directly on high-quality white punched cards, using either a Varityper Line Composer or an electric typewriter, as in Figure 1. After all the cards for an index volume are typed and arranged in the desired order,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

the cards are run through a Fotolist, Listomatic, or Compos-O-Line sequential card camera for high-speed exposure of a one-column-wide negative. If the printed page is to have more than one column, two or more strips of negatives are mounted side by side to make up a full-page negative for use in exposing an offset plate for printing.

The cameras are generally set to photograph one typed line per card. Only the Listomatic camera will reset its aperture automatically to expose two or three lines on a card in response to an appropriately positioned key punch. This automatic control feature is highly desirable to expedite the handling of index entries that run over one line long. Unfortunately, however, the variable aperture has been a major cause of machine malfunctions when the machine is operating close to its rated speed of 13,800 cards per hour.

With the Fotolist and Compos-O-Line cameras, the aperture can be set for a desired number of lines beforehand but cannot be changed automatically during a run. The fixed aperture and a somewhat slower rated speed of 7,200 cards per hour give essentially troublefree operation.

With typing on cards, the typist can work at her maximum speed without fear of being penalized for mistakes. When a mistake does occur, she simply removes the card, inserts another, and retypes the line with essentially no loss of typing rhythm. The knowledge that mistakes are so easily corrected increases the accuracy and volume of typing output at all levels of typing skill.

When index entries are typed on cards, there is usually no attempt to line up the right-hand margin. This lack of justification actually improves readability, even though the ragged right-hand margins may bother purists who have become accustomed to the traditional printed columns. When an entry terminates in numerical data, tab-

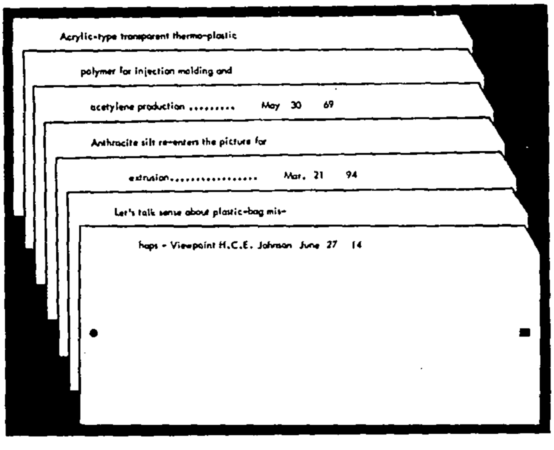


FIG. 1. Example of index typed one line at a time on IBM punched cards having special registry holes in columns 1 and 80 for use in Varityper Line Composer and Fotolist sequential card camera.

Cotton bleaching		
Better cotton bleaching, Combine chlorine and peroxide for	Mid-July	108
Bleaching, dyeing, finishing – modernization 1959..........	Oct.	77
Bleaching – modernization 1959	Oct.	71
Bleaching, What's new in	Nov.	97
Caustic boil, Per-acids replace	July	64
Continuous bleaching, Buck sour speeds	Mar.	70
Continuous bleaching with silicate-free peroxide	Mid-July	109
Cotton strength, How pH of bleach bath affects	Mid-July	111
Hydrogen-peroxide bleaching..........	Mid-July	110
Modern bleaching methods: a comparison	Mid-July	110
Peroxide behavior studied	May	70
Preparation & bleaching (a TEXTILE WORLD refresher) – 1. Apr. 97; 2. May 83; 3. June 82; *4. July 73; 5. Aug. 74;* 6. Sept. 131	Dec.	85
Processing in Lowell Bleachery South, A look at..........	May	78
Soda bleach liquors, Preparation of	Mid-July	107
Cotton blending		
Cotton Research Clinic, Blending and carding highlight..........	July	45

FIG. 2. Portion of annual index for *Textile World*, showing how use of tab key for page numbers at the end of each entry gives the effect of justification without the extra cost of retyping.

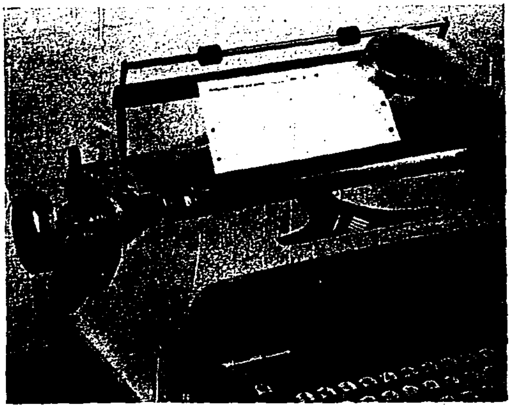


FIG. 3. Method of removing typed card from new IBM card-holding platen. Two positioning pegs, visible below lower corners of card, mate with extra pair of holes punched in card near these corners, when operator places card in metal strip and rolls platen backward to pull card down.

bing is often used to give the effect of lineup at the right, as in Figure 2.

The new IBM card-holding platen shown in Figure 3 now makes it possible to position the cards accurately in an electric typewriter and increase typing speeds up to 20% above those obtained with a Varityper Line Composer. The platen has two positioning pegs that are designed to mate with an additional pair of round and rectangular holes punched into a Fotolist card. These holes are low enough so the card can seat itself on the pegs as the operator rolls back the platen to pull the card down into the machine. With this new platen and the four-hole cards, accuracy is such that a card can be typed, taken out, put back in the typewriter and the entry typed over again without the retyping being noticeable.

• Examples of Camera-produced Indexes

The sequential card camera technique is currently being used successfully by the National Library of

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Medicine for producing *Index Medicus* (3) and by the Atomic Energy Commission in Oak Ridge for producing the index pages for *Nuclear Science Abstracts* (4).

At Chemical Abstracts, one of the largest index-publishing operations in the world, use of a Fotolist camera with Varityper Line Composers is at least as economical as conventional hot-metal composition for the alphabetical index to its abstracts (5). Entries are typed, proofed, and corrected concurrently with indexing. The index can then go to press within a day or two after the last entry is made. Chemical Abstracts plans to use the repro card technique first for its annual index. The cards can then be saved for use in later production of cumulative multi-year indexes.

An outstanding example of an annotated index, produced by typing cards with an IBM typewriter having a modified card-holding platen and running the cards through a Listomatic card camera, is the *Index Handbook of Cardiovascular Agents*. This was prepared under the direction of Dr. Isaac Welt (6, 7, 8). Each entry consists of one or more sentences describing in detail the content of the article as it pertains to cardiovascular agents, as in Figure 4. Abbreviations are used extensively to save space and thereby cut costs.

Three McGraw-Hill publications (*Electronics*, *Nucleonics*, and *Textile World*) are now using card cameras for producing annual indexes. With repro copy on IBM cards, index entries for each issue can be typed, proofread, and corrected throughout the year. Only the entries for the last issue to go in the index need to be done under the pressure of a deadline.

• Computers Speed Index Production

A computer can be used in place of punched-card sorting and processing machines to speed the production of indexes from entries received in random order on punched cards, punched paper tape, or Unityper magnetic tape. With computer processing of accurate input data, the output will be perfectly accurate even though a large number of automatic sorting and editing operations are called for.

← Venom, rattlesnake prod., when inject. i.v. in anesth. dogs.	02842
← Veratraldehyde prod., upon admin. in Ringer's sol. to isol. Helix aspersa heart. (Fr)	03660
← Veratrine prod., in cat upon s.c. admin. (Ger)	07796
← Veratrum viride, glycoside abol., in many of 32 pts. with circ. insuffic. upon i.v. or p.o. admin. (Rus)	08100
← Vinyl acetate prod., foll. inject. in rt. coron. art. of dogs.	03593
← Vitacamphor regulated, in isol. heart prep. of Entosphenus japonicus. (Jap)	13010
← Xenon did not prod., in pts. upon inhal. as mixt. with oxygen to ind. anesth.	10596
← Zinc hydroxide did not prod., upon i. ca. admin. through intact chest wall in anesth. dogs.	06219
ARSENIC	
→ Blood pressure, decr., foll. by death in few hours. (Fr)	09083
→ Capillary BF, decr., by vasoconstr. and damage of art. wall. (Ger)	05973
→ Cardiac arrest, prod. fatal, in few hours after incr. in HR and decr. in BP. (Fr)	09083

FIG. 4. Portion of subject index from *Index Handbook of Cardiovascular Agents*, showing document numbers lined up at right. Arrow symbols are used to indicate direction of interaction between subject heading and first words of index entry.

If computer output tapes could be converted automatically to repro copy by an errorless composing machine, there would be no need for further proofreading. A Linotype or other hot-metal typesetting machine does not meet these accuracy requirements, even though it can be controlled by punched paper Teletypesetting tape. The reason for this is an inherent residual mechanical error of about 2 per cent even in the best hot-metal slug-casting machines. These machine malfunctions mean that copy must be proofread, because an average of one slug in fifty will have errors that require resetting. The machine error has been inconsequential up to now because it is far overshadowed by human composition errors that can get as high as 20 per cent.

Error-free copy can be obtained from a high-speed printer fed by the output of a computer, but this has three serious faults for published indexes: (1) The printing is all capital letters, which are hard to read; (2) all characters are uniformly spaced, giving a low character count per line; (3) the number of special characters is limited, making it difficult or impossible to get desired special symbols.

One high-speed printer breakthrough has been introduced by IBM. Their new type 1403 high-speed printer can be built on special order to give both caps and lower case, though still with uniform spacing and a limited number of special characters. A drawback for some index publishers is the need for a 1401 or other computer to control this printer.

A computer can easily be programmed to produce punched paper tapes for a tape-controlled typewriter such as a Flexowriter, but this also has several drawbacks. First, the equipment for converting magnetic tape to punched paper tape is expensive. Second, the typewriter is limited in speed even when driven by tape. A large number of typewriters would be needed to produce repro copy for a large index in a reasonable time. Third, typewriters can have only one type face at a time. A Varityper Line Composer that can be driven by punched tape or punched cards and give a choice of type faces is still in the laboratory development stage, and would be subject to the same speed limitations as a typewriter.

A hopeful solution to this problem of getting accurate and readable computer-generated repro copy is based on the use of a tape-controlled photo-composition machine. Punched paper tapes for driving the Photon machine are already being produced by an IBM type 709 computer and a tape converter on an experimental basis, using programs produced by Prof. Michael Barnett and Kalon Kelly of MIT (9). A major advantage of this approach is the large number of type faces and special characters available on one type disk for index composition. Plans are under way to use the computer-Photon combination of equipment to produce a *Union List of Communist Chinese Serials*, under the direction of Ryburn M. Ross, Associate Director of Libraries for MIT.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

The chief drawbacks of photo-composition machines today are slowness (about 10 characters per second), relatively high cost in comparison with hot-metal composition, and high down time due to malfunctions. There is some hope that a 100-character-per-second tape-controlled Photon, now under development, will be more attractive economically and still give perfect composition from punched tape.

• Examples of Computer-produced Indexes

Perhaps the best-promoted example of a computer-produced index is the permuted-title index (10, 11). Here the entire index-producing operation is performed automatically once the titles and related bibliographic data have been keypunched for use as input to the computer. In this type of index, each significant word of the article title is used in turn with its context as an index term. Title recirculation, also called snapback, is usually used to place as much of the title as possible in the available space, as shown in Figure 5.

Permuted-title indexing serves primarily for current awareness, but also has some merit for searching in fields where more conventional indexes do not exist or are not sufficiently current. The chief advantage is elimination of the creative work of human indexers, since computer production processes can be applied to other types of indexes as well. Permuted-title indexes are reproduced at present by offset printing, working directly from the all-caps copy produced by the high-speed printer. The IBM version of this index is called KWIC, an acronym for Key Word In Context.

A number of other organizations are either studying or using the permuted-title technique. The initial success of the published permuted-title indexes has encouraged some to consider adding a small amount of editorial work to further improve the value of the index. This can be done by rewriting the title to get in a larger number of significant words (needed badly for articles in trade journals), or by adding selected key words to the title as is done in the *KWIC Index to Neurochemistry* (12).

The first permuted-title index to be published regularly was *Chemical Titles*, launched in January 1961 on a semimonthly basis by the American Chemical Society (13, 14). Although new publications generally require a number of years to get out of the red, current income from almost 3,000 subscribers at $50 per year has already reached the point where it covers production and distribution costs. A National Science Foundation grant, permitting distribution of over 100,000 copies of sample issues in 1960 for promotional purposes, is largely responsible for this early success. An average issue today covers about 2,700 documents and runs over 100 pages.

A sample issue of *Chemical Patents*, an index of permuted keywords of patent titles, has also been produced by Chemical Abstracts. Here the majority of patent titles had to be rewritten to make the titles reveal rather than conceal the content of the patents.

A permuted-title index is currently being used by Bell Laboratories for internal reports (15). *Meteorological Titles* is another example that is already on a regular published schedule. Biological Abstracts launched its *Biochemical Title Index* in January 1962 on a monthly basis as a current awareness service, using IBM's KWIC format (16).

The ultimate in permutation for indexing is a published concordance. This indexes every word in a document, with its context, hence is of value only when the searcher seeks usages of words. There are now several examples of computer-produced concordances in print, including those for the Bible, the Dead Sea Scrolls, and the writings of St. Thomas Aquinas.

Another type of computer-processed index, prepared under the direction of Y. S. Touloukian of Purdue, has been published as a three-book set by McGraw-Hill under the title, *Retrieval Guide to Thermophysical Properties Research Literature* (17). The goal of this index is to locate articles that deal with specific properties of a specific substance.

The importance of typographic experimentation in connection with index publishing is dramatically shown in Figures 6 and 7. The original format of the Thermophysical Properties index required two full pages as produced on an ordinary electric typewriter, as in Figure 6. Experimentation with a Varityper Line Composer showed that the same data could be printed on one page without loss of readability through use of a compressed type face, proportional spacing, and special subscript characters, as in Figure 7. For numerical data alone, in another section of this index, direct use of computer printout was entirely satisfactory.

The high speed of computer processing made it possible to produce and publish at convention time an index to over 2,500 abstracts of papers presented at the 1960 meeting of the Federation of American Societies for Experimental Biology. The index was a byproduct of an experiment in planning the program of a large convention with the aid of a computer, carried out under the supervision of Claire Schultz (18). The alphabetic index was retyped from the high-speed printer output for the

```
JONE-RA61-DSG  JONES RA
                 DETERMINATION OF SULFUR IN GASOLINE BY X-RAY EMISSION
                   SPECTROGRAPHY.=
                 ANAL. CHEM.. 33. 71-3 (1961).


MISSION SPECTROGRAPHY.= DETERMINATION OF SULFUR IN GASOLINE   JONE-RA61-DSG
APHY.= DETERMINATION OF SULFUR IN GASOLINE BY X-RAY EMISSION  JONE-RA61-DSG
ERMINATION OF SULFUR IN GASOLINE BY X-RAY EMISSION SPECTROGR  JONE-RA61-DSG
F SULFUR IN GASOLINE BY X-RAY EMISSION SPECTROGRAPHY.= DETER  JONE-RA61-DSG
UR IN GASOLINE BY X-RAY EMISSION SPECTROGRAPHY.= DETERMINATI  JONE-RA61-DSG
OLINE BY X-RAY EMISSION SPECTROGRAPHY.= DETERMINATION OF SUL  JONE-RA61-DSG


JONES RA        JONE-RA61-DSG
```

FIG. 5. Example of *Chemical Titles* entry produced by computer for one document when bibliographic data at top is keypunched as input. Computer fills entire space available for title by transferring unprinted remainder of title at end of line to the start of that line when room is available at the start. End of article title is indicated by equality sign. Entries are alphabetized by the words *determination, sulfur, gasoline, x-ray, emission, and spectrography,* which line up near center.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

560

CLASS 516

Alphabetized Formula	Molecular Formula	Substance No.
$C_{14}H_9ClO_3S$	$ClC_{14}H_8SO_3H$	516-0391
$C_{14}H_9ClO_3S + ClH$	$ClC_{14}H_8SO_3H + HCl$	516-0393
$C_{14}H_{10}HgO_{12}S_2$	$(C_6H_3OHSO_2OHCOO)_2Hg$	516-1061
$C_{14}H_{10}O_9$		516-0216
$C_{14}H_{27}KO_2$	$CH_3(CH_2)_{12}COOK$	516-1226
$(C_{14}H_{27}KO_2 + ClNa + HKO)$	$CH_3(CH_2)_{12}COOK + NaCl + KOH) \rightarrow CH_3(CH_2)_{12}COOK$	516-0335
$\rightarrow C_{14}H_{27}KO_2$		516-0335
$C_{14}H_{23}O_3$	$C_{11}H_{23}COO(CH_2CH_2O)_nH$	516-1032
$C_{14}H_{29}NO_2$		516-1231

FIG. 6. Format of main index used in the first published volume of the *Thermophysical Properties Index*, as prepared on ordinary typewriter, with the listing extending across two full pages.

581

CLASS 516

Substance No.	Name	Property Code	
516-0391	9-Chlorophenanthrene 3-sulfonic acid	3	
516-0393	9-Chlorophenanthrene 3-sulfonic acid + Hydrogen chloride	3	
516-1061	Mercuric sulfosalicylate	3	
516-0216	Tannin	3	5
516-1226	Potassium myristate	3	
516-0335	[Potassium hydroxide(3) + Potassium myristate(1) + Sodium chloride(2)]		
516-0335	→ Potassium myristate		5
516-1032	Polyethylene glycol monocaprate	3	
516-1231	Hexanolamine caprylate	3	

FIG. 7. Proposed new format for main section of *Thermophysical Properties Index*, showing how compressed characters and proportional spacing permit cutting width of listing in half so it will go on a single page. The property code numerals at the right tell which of seven possible properties of the named material are covered in the reference.

1960 edition, but the computer printout was used directly for repro copy for the 1961 edition (19).

Index Chemicus also uses complete Univac computer processing to aid in achieving its goal of 30 to 45 days between receipt and publication (20). The molecular-structure diagrams used in this index permit browsing both for current awareness and random searching. A newly developed computer program (21) is now being used to translate the name of a chemical compound automatically to its correct molecular formula.

• New Approaches to Index Publishing

Each issue of *Factory* magazine contains an index page in which article titles are grouped under subject headings and further identified by a numerical code, as in Figure 8. This is a classification-type index somewhat like the Dewey Decimal system (22). Cross-reference numbers are given in parentheses after an entry where appropriate.

The reader is instructed to cut up the index items, paste them on 3x5 cards, and file by code number. To cross-index, the user copies the information on a second card and files it by its secondary code. Here, then, is a do-it-yourself version of a document filing system. Reader surveys over the past three years have indicated sufficient use of this indexing service to justify its continuation.

This year the American Institute of Chemical Engineers launched its plan for making authors do their own indexing and abstracting of articles for *Chemical Engineering Progress* (23). Each author is expected to use as a guide the recently published *AIChE Thesaurus* of approved key words, links, and role indicators for concept-coordination indexing (24). The resulting index terms are published with the abstract in each article, for use by readers in setting up their own abstract-index card file, as in Figure 9. The user must transfer the published data to his own cards.

Although not a new approach, many information specialists feel that the citation indexes used by lawyers can be a valuable supplement to other types of existing indexes (25). A citation index enables a searcher to go from an earlier paper to all later papers that cite it as a reference. *Shepard's Citations* (26), published since 1873, is an example that serves the legal profession by citing each court case and listing all subsequent publications, court decisions, and other references relating to that case.

An exhaustive three-year study of citation indexes is now being made by Eugene Garfield for the National Institutes of Health. The goal will be publication of a citation index to the field of genetics, involving the processing of some ten million references in about 500 scientific journals.

A number of coordinate indexes are being published in book form for conventional manual use. In two examples, the semimonthly *Technical Translations* (27) and the *Quarterly Cumulative Index* for ASTIA's *Technical Abstract Bulletin* (28), only the main descriptors are used as subject headings for the published index. These are identified by asterisks in the list of descriptors that accompanies each abstract, as in Figure 10. The appearance of the resulting index is shown in Figure 11.

For your action index

1.0–1.4 Plant Maintenance and Engineering

MANAGEMENT AND PRACTICES (1.1)
1.106 Electrical troubleshooting (Nov. '60) p. 166 (3.201)
1.107 Memo-motion improves maintenance (Nov. '60) p. 84
1.107 Quality control in maintenance (Nov. '60) p. 102
1.107 Go straight on work measurement (Nov. '60) p. 162

BUILDING AND YARDS (1.2)
1.203 Epoxy for floors (Nov. '60) p. 170 (1.11)
1.209 Atlanta, city of contrasts (Nov. '60) p. 62
1.210 Stop explosion damage (Nov. '60) p. 88 (3.312)

SERVICES AND FACILITIES (1.3)
1.301 Make air conditioning pay off (Nov. '60) p. 94
1.301 Cool air computer (Nov. '60) p. 162 (1.403)
1.302 Cleaning up the air (Nov. '60) p. 168 (1.111)

2.404 Platform for steel wire (Nov. '60) p. 171 (2.308)
2.404 Xmas tree rack (Nov. '60) p. 171 (2.409)
2.406 Custom-made skids for handling (Nov. '60) p. 172
2.409 Stretching storage space (Nov. '60) p. 168 (2.406)
2.410 New use for lawn tractors (Nov. '60) p. 165 (2.404)

3.0–3.4 Plant Management

ORGANIZATION AND POLICIES (3.1)
3.101 Replacement plan watches budget (Nov. '60) p. 164
3.107 Low-down on the election (Nov. '60) p. 171 (3.305)
3.107 Paper munitions for Mideast (Nov. '60) p. 198

PERSONNEL DEVELOPMENT (3.2)
3.201 They work production wonders (Nov. '60) p. 96 (3.405)
3.202 Seniority (Nov. '60) p. 89 (3.314)
3.204 Wife holding you back? (Nov. '60) p. 201 (3.206)
3.207 The ubiquitous manager (Nov. '60) p. 100 (3.305)
3.207 A snub for a smile (Nov. '60) p. 202 (3.303)

EMPLOYEE RELATIONS (3.3)
3.303 Robot brain panic button (Nov. '60) p. 166 (3.303)

FIG. 8. One of the few examples of a published classification-type index. This is used in each issue of *Factory* magazine, a McGraw-Hill publication, to help its readers establish their own filing system for articles.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A modified coordinate or Uniterm indexing system was used in producing a published index to research projects supported by the National Institutes of Health during the fiscal year 1961 (29). An average of 12 main terms, each having about five modifiers, is extracted for summary statements of projects and typed on IBM cards. After sorting, the cards are filed for current retrieval until needed for the preparation of photocopy by a Listomatic camera.

• Avenues for Further Research

There is still ample opportunity for coordinated research aimed at producing more and better indexes at lower cost. Possible avenues of approach for this research include the following:

1. Explore the possibility of standardizing the index format for various scientific fields, to facilitate interdisciplinary use and permit development of standard computer programs for processing indexes. This project could include the improvement of subject heading lists and thesauri in each field. This index standardization program is an important part of the BIOMEDIC information complex proposed by Dr. Richard Orr (30).

2. Explore the possibility of applying programmed teaching to indexing, with or without machines. Programmed aids for indexers can reduce training time, cut costs by speeding up actual indexing, and give greater indexing consistency among different indexers at different times. As one approach, each choice of an indexing term could place in front of the indexer a display of questions or possible additional indexing terms. These would be arranged to guide his thinking to the next logical choice of an indexing term. Rome Air Development Center and the National Library of Medicine are concerned with this same problem.

3. There is still need for an exhaustive across-the-board study of how indexes are actually used. This study should be coupled with a search for practical ways to get more user feedback, so that indexers are kept in touch with the current needs of searchers. One problem here is that of deducing what users need from their actions, since they are usually unable to tell specifically what they want. With computer-based information centers, the data needed for a user analysis can be obtained automatically as a byproduct of searching.

4. Explore techniques for speeding up the creative human effort required for conventional indexing, for use with computer processing to bring out these indexes much faster. Success here, combined with the use of computer processing, could eliminate the need for publishing temporary permuted-title indexes.

5. Study ways of improving the typography, readability, and ease of use of indexes. As one example here, explore the merit of user complaints about having to turn back and forth in some indexes to get complete bibliographic information, versus the increased cost of a larger index in which each entry provides complete information. Explore also the economic aspects of printing indexes in much smaller type and furnishing a magnifying glass with each volume. The chief savings here would be in plate-making, printing, and binding costs, because composition costs do not change appreciably with type size. At what point will the savings in printing costs offset the howls of indignant librarians?

a companion article by Kuehn and Davidson immediately following (*11*). However, before doing so, tests for stability of conditions within the unit are employed to determine if operation is stable. If not, computation of

established, but early prophetic estimates of direct benefits totalling ½ to 1¢/bbl. of crude oil processed (before taxes or computing costs) seem real-

Information Retrieval*

Key words: 1. Distillation-H, Separation-H, Controlling-H, Distillation-I, Separation-I, Petroleum-I, Crudes-I, Instrumentation-J, Computers-J, IBM 704-J.
Abstract: A computer control system developed by American Oil and IBM for a 140,000 bbl./day distillation unit at Whiting, Indiana, has undergone improvements and evaluation since initial operation in Oct., 1960. Frequent evaluaton of equipment performance, and determination of feed and product qualities, results in maximum yield of valuable products with minimum operating expense.
* For details on the use of these key words and the A.I.Ch.E. Information Retrieval program, see CEP, p. 55, May, 1961; p. 73, June, 1961.

FIG. 9. Example of index and abstract format now used with each article in *Chemical Engineering Progress*. Letters after key words designate role indicators.

Ordnance, Missiles, and Satellite Vehicles

TRACKING OF MISSILES AND SPACE VEHICLES: REVIEW OF SOVIET LITERATURE. Monthly rept. no. 9. 28 Feb 61 [9]p. 5 refs. AID rept. 61-22; AD-254 392.
Order from OTS or SLA $1.10 61-19625

DESCRIPTORS: *Satellite vehicles, *Guided missiles, Guided missile tracking systems, *Tracking, *Electromagnetic waves, Ions, Ionospheric disturbances, Radio astronomy, Bibliography, USSR.

Contents:
Electromagnetic problems
Ion clouds and ionosphere perturbations
Radio astronomy.

FIG. 10. Format used in *Technical Translations* (*OTS*) for listing descriptors. Asterisks designate descriptors that are also used as subject headings in the corresponding published index.

Subject	Page	Number
Electromagnetic waves	179	61-19625
	221	61-23180
	226	61-16098
		MDF B-185
Electromagnetism	220	61-19904
Electron microscopy	221	BISI-2043
Electron tubes	176	61-31116
Electronic equipment	176	61-21679
	196	ATS-46N51G
Electroplating	227	61-21751
Electroslag remelting	205	HB-5075

FIG. 11. Published index for *Technical Translations*, showing how several document numbers are listed under some of the descriptors. Descriptors assigned to documents in this index serve to provide greater depth of searching by computer.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

6. Improve the sequential card camera technique for producing indexes. A more reliable means of exposing a variable number of lines automatically under control of key punches is needed. Equally desirable is a low-cost proportional-spacing electric typewriter in which the type face can be changed as easily as on a Varityper. Still another needed feature is the ability to store an entire line in a typewriter, while proofreading it, then punch a button to get automatic retyping with justification as in Linotype machines.

7. Develop low-cost equipment for duplicating typed cards automatically, completely or in part, at higher speed and lower cost than can be achieved today with tape-controlled typewriters. With only one typing of data, typed index cards for all desired listings could then be created automatically without need for proofreading.

8. Continue the development of computer programs and character readers for completely automatic creation and printout of indexes. Although progress has been made by researchers (31, 32, 33, 34), we are still a long way from equalling the performance of human indexers.

9. Develop a high-speed printer capable of graphic arts quality, so as to equal the readability of conventionally printed indexes. The requirements here of proportional spacing, caps and lower case, two or more different fonts, and a variety of special characters, appear to point to an electronic solution. As yet, however, no cathode-ray type of printer has been demonstrated that will give the required resolution for printing small characters without losing sharp serifs or filling in letters.

10. Study the feasibility of publishing a master index to all of the world's material that merits indexing, with automatic cumulation at regular intervals and creation of specialized indexes as an automatic byproduct. The resulting computer magnetic tapes could then be released to information centers that use computers for searches, thereby eliminating costly duplication of indexing and keypunching effort.

In conclusion, the status of published indexes today can be summarized by these five points:

1. This country is now publishing some of the finest indexes in the world.
2. Most index-publishing operations have already adopted some forms of mechanization to speed publication and cut costs.
3. Many more specialized indexes are needed to cover existing gaps in indexing.
4. More research should be aimed at improving the usefulness of published indexes to librarians and scientists.
5. More money is needed to pay for all this.

References

1. Cleverdon, C. W. September 1960. Report on the first stage of an investigation into the comparative efficiency of indexing systems. ASLIB Cranfield Research Project. The College of Aeronautics, Cranfield, England.
2. Cleverdon, C. W. November 1960. Interim report on the test programme of an investigation into the comparative study of indexing systems. ASLIB Cranfield Research Project.
3. The National Library of Medicine Index Mechanization Project. National Library of Medicine, Washington, D. C., 1961. Part 2, *Bull. Med. Lib. Assoc.* **49**(1): 1–96.
4. Day, M. S., and I. Lebow. 1960. New indexing pattern for *Nuclear Science Abstracts. Am. Document.* **11**(2): 120–127.
5. Dyson, G. M. 1961. Current research at Chemical Abstracts, *J. Chem. Doc.* **1**(1): 24–28.
6. Welt, I. D., 1961. Guide to the world literature on cardiovascular agents. *Med. Doc.*, **5**: 9–10.
7. MacMillan, J. T., and I. D. Welt. 1961. A study of indexing procedures in a limited area of the medical sciences. *Am. Doc.* **12**(1): 27–31.
8. Welt, I. D. *Index-Handbook of Cardiovascular Agents*, **2** (1951–1955), 1,568 pages (Washington, National Academy of Science–National Research Council, 1960).
9. News Release on computer programming by M. P. Barnett and K. L. Kelly for operating Photon typesetting machine, Office of Public Relations, Massachusetts Institute of Technology, Cambridge, Mass.
10. Citron, J., L. Hart, and H. Ohlman. A permutation index to the Preprints of the International Conference on Scientific Information. System Development Corp., SP-44, rev. ed., December 1959, originally published November 1958.
11. Luhn, H. P. Keyword in context index for technical literature (KWIC index). IBM Advanced Systems Development Division, Report RC 127, August 1959, *Am. Document.* **11**(4), 288–295.
12. *KWIC Index to Neurochemistry.* Mimosa Frenk Foundation, Amsterdam, The Netherlands, & IBM.
13. "Chemical literature gets a quicker index." *Chem. & Eng. News* **38**(14): 27–28.
14. *Chemical Titles.* Published by The American Chemical Society, Washington, D. C.
15. Kennedy, R. A. Mechanized title word indexing of internal reports. Presented at the Third Institute of Information Storage and Retrieval, The American University, Washington, D. C., February 13-17, 1961.
16. *Biochemical Title Index.* Sample issue published July 1961 by Biological Abstracts, 3815 Walnut St., Phila., Pa.
17. Touloukian, Y. S. 1961. *Retrieval Guide to Thermophysical Properties Research Literature* **1** (3 books). McGraw-Hill Book Co., New York.
18. Schultz, C. K., and C. A. Shepherd. The 1960 Federation meeting: Scheduling a meeting and preparing an index by computer. *Federation Proc.* **19**(2): 682–699.
19. *Federation Proceedings.* Federation of American Societies for Experimental Biology **20**(1): Part I—Abstracts, Author Index and Subject Index.
20. *Index Chemicus.* Published by Institute for Scientific Information, Phila., Pa.
21. Garfield, E. 1961. An algorithm for translating chemical names to molecular formulas (Doctoral Thesis). Institute for Scientific Information, Phila., Pa.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

22. Getting even more out of Factory; classification system for Factory's problem-solving guide. *Factory Management and Maintenance* **115**(2): 104–105.
23. HOLM, B. E. 1961. Information retrieval—a solution. *Chem. Eng. Progress* **57**(6): 73–78.
24. *Chemical Engineering Thesaurus*. 1961. American Institute of Chemical Engineers, New York.
25. GARFIELD, E. Citation indexes for science. *Science* **122**(3159): 108–111.
26. *Shepard's Citations*. Published by Shepard's Citations, Inc., Colorado Springs, Colo.
27. *Technical Translations*. Published by the Office of Technical Services, Washington, D. C.
28. *Quarterly Cumulative Index to Technical Abstract Bulletin*. Published by Armed Services Technical Information Agency, Arlington, Va.
29. *Research Grants Index*. Sample issue for Fiscal Year 1961. National Institutes of Health, Bethesda, Md.
30. ORR, R., BIOMEDIC (Biomedical Information Complex). Institute for the Advancement of Medical Communication, New York City.
31. BAXENDALE, P. B. Machine-made index for technical literature—an experiment. *IBM J.* (10): 354–361.
32. SWANSON, D. R. 1960. Searching natural language text by computer. *Science* **132**(3434): 1099–1104.
33. MARON, M. E. Automatic indexing: an experimental inquiry. *J. Assoc. Comp. Mach.*, **8**(3): 404–417.
34. EDMUNDSON, H. P., and R. E. WYLLYS. Automatic abstracting and indexing—study and recommendations. *Communications of the Assoc. for Comp. Mach.* **4**(5): 226–234.

DISCUSSION

Moderator Adams: By pre-arrangement, three of the panelists have drawn assignments for extending Mr. Markus's paper. Their remarks will be on specialized aspects.

Dr. Eugene Garfield, President of the Institute of Scientific Information, has been asked by the panel to comment on the psychological factors involved in the display features of a published index.

Dr. Eugene Garfield: The proper selection of typographical style is an important ingredient for consumer acceptance of printed publications. Unfortunately there is no quantitative measure of this extremely subjective consideration—the esthetic and psychological reasons why we accept one format and reject the other. Type styles change and so do art forms. We know very little about the motivation for this. Publishers can only follow their intuitions on these matters.

The brevity of this talk does not allow me to mention many relevant studies that have been made along these lines. It is my intent merely to reiterate the existence of the problem of proper packaging of the printed word. I will use figures to illustrate my points. For the figures I have chosen three kinds of material related to indexes: display of contents pages (Fig. 1), high-speed printer output from a computer (Fig. 2) and graphic versus verbal abstracts (Fig. 3).

Those readers who are acquainted with printing techniques will recognize that the variations in the different formats represent manipulation of such things as style of type, size of type, camera reduction of the prepared copy, column width, number of columns, leading (spacing), use of attention devices such as dotted lines, arrows, stars, etc., alterations in type style within the piece, alteration of upper and lower case, and alteration of the order of bibliographic elements.

The advent of computer actuated type-setting and photo-setting devices makes it all the more important and advantageous to focus attention on the problem of finding the most effective formats for the composition of the printed page.

Moderator Adams: For the benefit of those who could not be present and did not hear the discussion yesterday afternoon, Mr. Herbert Ohlman, of IBM, has some comments.

Mr. Herbert Ohlman: In recent years, several cold-type or photocomposing systems have reached the market. However, they are either key-driven, and therefore tied to the accuracy and speed of their operators, or paper-tape driven. An example of the latter is Mergenthaler's Linofilm, which could be used to set copy for a 300-page book in about 12 hours. In the future, such machines may accomplish this work in as little as one hour.

However, the output requirements of electronic data-processing systems are at least a thousand times greater than Linofilm could handle. These radically faster systems have spurred the development of radically different composing machines, the so-called line-a-time or "high-speed" printers. Most available high-speed printers work on an electro-mechanical principle, with their type characters (on wheels or chains) being struck by solenoid-driven hammers; some create characters on the spot from a 5 by 7 matrix of stiff wires.

The up to 1,000 lines per minute speeds that these devices attain are mainly the result of paralleling type characters, so that an entire line of from 60 to 130 characters is printed at one stroke. Our hypothetical 300-page book could be "composed" in this manner in only 10½ minutes, but it would not look much like a book, because these devices are capable of printing only a very limited number of characters—usually between 40 and 50 (numerals, upper case letters, and a few special symbols).

However, fast as these machines are, they have just about reached their limit; they are inherently mechanical, and attempting to get a greater variety of characters must be paid for in terms of slower speeds, or poorer quality, or both. For on-line use, data-processing machines will require output devices to compose at upwards of 50,000 lines per minute—a book every 12 seconds. The only way to attain such speeds is to abandon ink and hammers for light and electrons. Of course, we still have to pay a price, and this time it will be in cost and complexity.

Three companies produce complete systems capable of from 5 to 20M lpm for upwards of $300,000; the rest produce the character-generating portion or the character-printing portion, but currently not both.

The character-generating systems are based without exception on cathode ray tubes, differing principally in

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

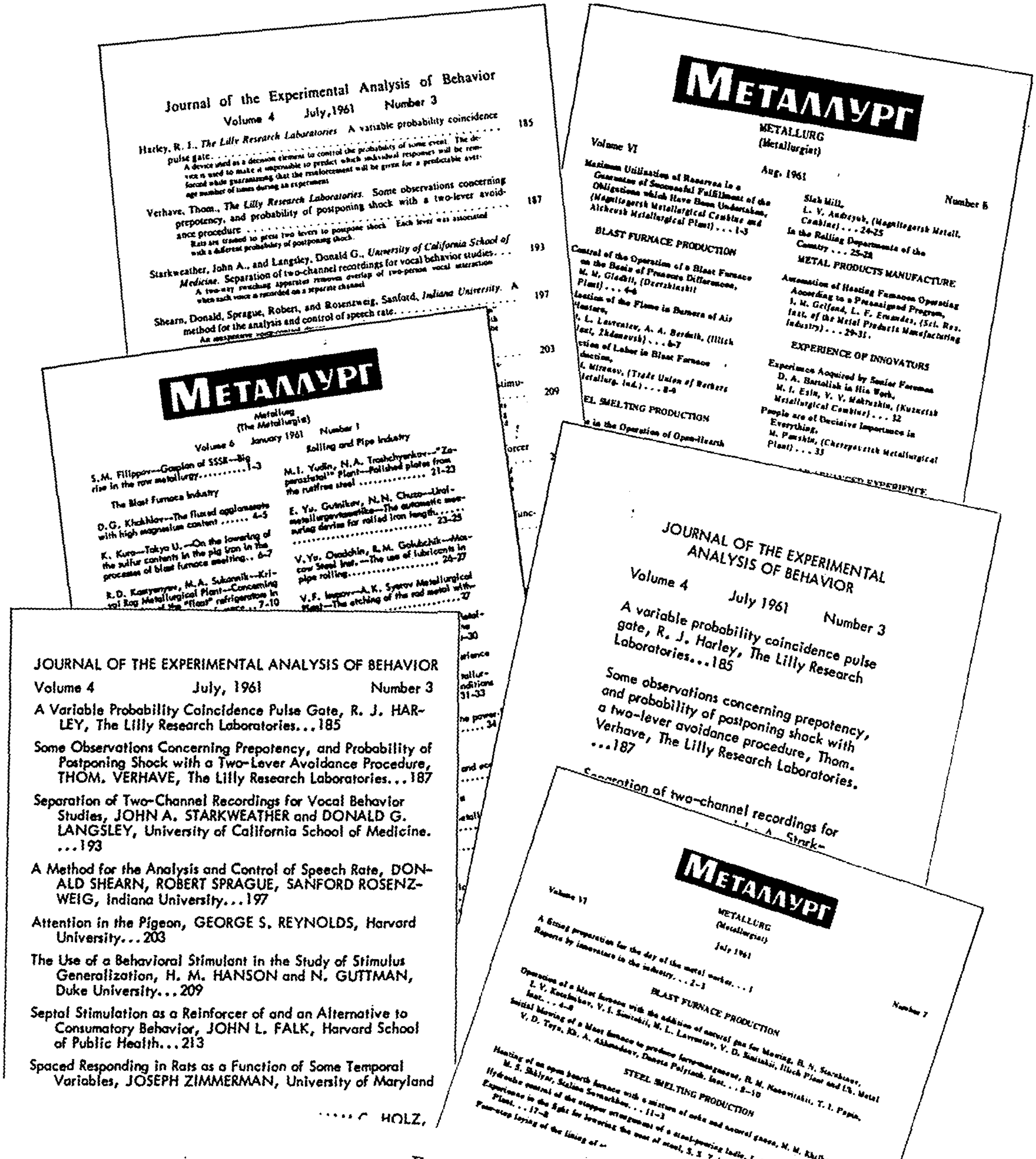


Fig. 1.

whether the characters are formed within the tube, or externally. Several are based on the shaped-beam tube, in which a metal stencil having an 8 by 8 matrix of characters literally "shapes" a beam of electrons; characters thus formed are formatted upon a fluorescent screen by deflection plates. The videograph monoscope tube varies this by utilizing characters printed on an aluminum screen, a video signal being generated by secondary emission of electrons from the screen. In the vidiac and Transdata systems characters are generated by solid-state electronics for separate CRT display. In these systems speed can be traded for practically unlimited character sets. In Videofile, a special TV camera converts characters or analog data to a video signal which is stored on magnetic tape. In the DACOM system, a conventional flying-spot cathode ray tube is used to scan a matrix of characters

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

on a glass plate with light; the modulated spot is then converted into a video signal by a matrix of photo-cells, which signal is in turn fed to another cathode ray tube to provide the display. In DACOM, a split-beam optical system permits format slides to be superimposed upon the printing medium (film). Most of these systems claim a speed of about 20,000 characters a second, but some are eventually shooting for 500,000—a book a second!

For printing, one of these character generators is coupled to either a conventional lens-and-film system or an electrostatic system. Most of the film systems employ 16 or 35-millimeter microfilm, which must be wet-processed and enlarged before copy can be run off. The SC-5000 uses a xerographic drum for on-line processing of full-size copy, and Videofile an Electro-fax printer Videograph incorporates a special electrostatic printing tube. This tube has wires embedded at right angles to its face, enabling impinging electrons to form an electrostatic latent image directly on a web of dielectrically coated paper. In these systems, analog signals, such as from television and facsimile generators, can be reproduced as easily as characters.

Once these new devices become capable of forming a full set of upper- and lower-case characters, and conventional punctuation marks, we may expect to see them employed on large-scale indexing projects, perhaps before they are used for continuous text composition. This is so because it is already possible to generate some types of index entries automatically, and store and cumulate them on magnetic tape. Such tapes work at speeds beginning around 20,000 characters a second, and may approach 200,000, thus providing the possibility of unbuffered coupling with character generators.

However, it is difficult to realize what such speeds mean. The entire 1957 book composition requirements of 800 U. S. publishers could have been met by one 20,000 line-a-minute machine in less than one work week. The world's largest published index, the cumulated *Index Medicus*, which provides author and subject access to 125,000 articles every year, would take only an hour of its time. Even were we to attempt to index all scientific and technical fields to the depth and breadth of the *Index Medicus*, we would be finished in one day. What do we feed the beast then?

Perhaps we should reconsider slower, but more versatile devices for composing indexes for publication. What we require is a device that can create a visually attractive page for a competitive price at reasonable speed. It should provide a full variety of type fonts, the gamut of colors, and automatic control of placement and shape, as well as the ability to compose analog matter, such as graphs and pictures, from digital inputs. The practical application of such displays as Chromatron photochromic dyes, coupled with such reproduction means as thermoplastic recording or smoke printing will usher in an era of true electronic printing in all the flexibility we have come to expect in printed communication.

Moderator Adams: The last of the formally assigned panelists is Dr. Robert Ledley, President of the National Biomedical Research Foundation. His comments will be about new generic and correlative indexes.

Dr. Ledley: My comments will be concerned with

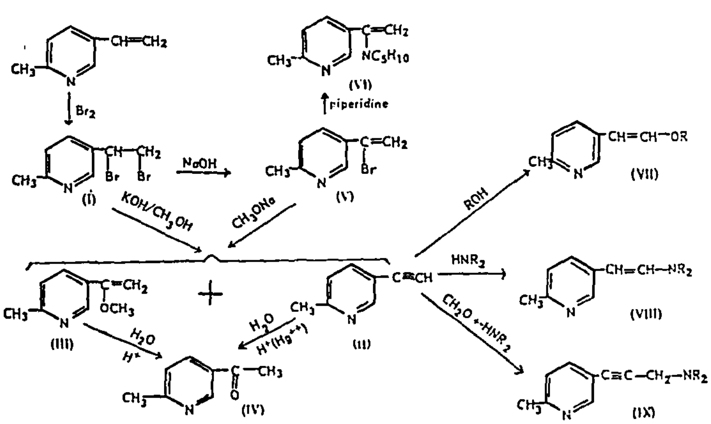
```
ACETYLCHOLINE
        *FORM., EFF. OF FOLLICULIN                  27080
    DETERMINATION
        *IN BLOOD P VISCERA AFTER DEATH BY
          VIOLENCE                                  25355
    DERIVATIVES
        REVERSED CARBOXYL ANALOGUE - DERIV.         28909
    EFFECTS
        ON RESP. IN ATROPINIZED - NON/ATRO/
          PINIZED CATS                              24709
        ON TRANSM. OF NERVE IMPULSE                 27470
        *UTERUS CONTRACTIONS, IN PRESENCE OF
          FOLLICULIN                                24796
    THERAPEUTIC USE
        *RAYNAUD S DIS., WITH PAPAVERINE            31319
ACETYLCHOLINESTERASE  SEE  CHOLINESTERASE
ACETYLSALICYLIC ACID
    DERIVATIVES
        *IN PREV. OF DEGLUTITION PAIN               29762
    EFFECTS
        *ON EOSINOPHIL COUNT                        26157
    TOXICITY
        CAUSING GASTROINTESTINAL HEMORRH.           28703
ACHONDROPLASIA
        OF JAWS, COMPARATIVE ANAT.                  27178
ACID-BASE EQUILIBRIUM
   SEE ALSO ACIDODIS,HYDROGEN-ION CONCENTRA-
          TION
        OXIDATION-REDUCTION-ACID-BASE CON-
          TINUUM                                    26711
        *IN PEPTIC ULCER                            32529
    DETERMINATION
        *                                           25982
        *                                           26014
ACID NUCLEOTIDASE  SEE  NUCLEASES
ACIDOSIS
        HYPERCHLOREMIC, RENAL                       30860
    EXPERIMENTAL
        *EFF. ON GENITALIA & ADRENALS OF
          STERILE RABBITS                           29129
    THERAPY
        ACTH & CORTISONE, IN CATTLE                 31531
ACIDS
   SEE ALSO  FATTY ACIDS, NUCLEIC ACIDS
        ACONITIC, BY-PRODUCT IN MANUFACTURE
          OF SUGAR                                  24794
ACNE
        CONGLOBATA, RELATION TO DISSECTING
          CELLULITIS & HIDRDENITIS                  24395
    BLOOD IN
        SERUM LIPIDS                                28598
    ETIOLOGY AND PATHOGENESIS
        RADIOIODINE THER. OF HYPERTHYROIDISM28437
ACNE ROSACEA  SEE ROSACEA
```

```
LASTIC FLUIDS.* SIMILAR SOLUTIONS.                                    SCHOWR-60-ABT
LESCENCE OF POLYSTYRENE SOLUTIONS.                                    DEBYEP-60-COP
LUTIONS.  LOW-MOLECULAR SOLUTIONS.                                    NYVLTJ-60-TNS
MAGNESIUM PHTHALOCYANIN SOLUTIONS.                                    AKIMIA-60-ISE
H FISSION PRODUCT WASTE SOLUTIONS.                                    GRAYPR-60-SCF
N OF ACIDS FROM AQUEOUS SOLUTIONS.                                    IVANLS-60-MAE
OOD FROM DILUTE AQUEOUS SOLUTIONS.                                    KLEITN-60-ASW
RROCYANIDE FROM AQUEOUS SOLUTIONS.                                    SOCHVG-60-ISI
 OF METALS FROM COMPLEX SOLUTIONS.                                    UKSHEA-60-TEM
 OF PROTEINS IN AQUEOUS SOLUTIONS.                                    KHOMNE-60-SPA
TED SEA WATER AND OTHER SOLUTIONS.                                    MACLRA-60-FSE
T SCATTERING OF POLYMER SOLUTIONS.                                    KOBOYY-60-EPL
UCTIVITY OF MOLTEN SALT SOLUTIONS.                                    LANTMF-60-EMS
UTRAL OR NEARLY NEUTRAL SOLUTIONS.                                    KRISHB-59-KMR
VER BROMIDE AND SULFIDE SOLUTIONS.                                    SCHAGM-60-MOT
 IN POTASSIUM HYDROXIDE SOLUTIONS.  (GER.)                            POSPIJ-60-ACO
TER OF ALKALI HYDROXIDE SOLUTIONS.  (GER.)                            CUTAF -60-MBM
ONS OF FREE RADICALS IN SOLUTIONS.  DESTRUCTION MECHANISM OF          DOLGBA-60-RFR
ECHANICAL PROPERTIES OF SOLUTIONS.  GELS AND COAGULATES (COA          PCHEVA-60-SMP
EORY OF NON-ELECTROLYTE SOLUTIONS.  LOW-MOLECULAR SOLUTIONS.          NYVLTJ-60-TNS
HROMIUM IN ACID AQUEOUS SOLUTIONS.  POTENTIOMETRIC TITRATION          SHUTSG-60-MCA
 COEFFICIENT OF POLYMER SOLUTIONS, AS DETERMINED BY THE LIGH          ESKIVE-60-DVC
              CELLULOSE SOLVENT CADOXEN, A PREPARATION, AND           HENLEO-60-CSC
LATO CHROMIUM(III) ION- SOLVENT DEUTERIUM ISOTOPE EFFECT.             KRISKV-60-SRO
  ENOLATES.  CATION AND SOLVENT EFFECTS IN THE ALKYLATION OF          ZOOKHD-60-ECS
 HALOGEN SUBSTITUENTS.  SOLVENT EFFECTS ON THE PRINCIPAL ELE          SCHUWM-60-HSS
TUENTS.  P- HALOGEN AND SOLVENT EFFECTS ON THE PRINCIPAL ELE          SCHUWM-60-HSP
OF SOME HEAVY METALS BY SOLVENT EXTRACTION AND BY PAPER CHRO          WEIDNG-60-SSH
ETERMINATION OF IRON BY SOLVENT EXTRACTION AS ACETYLACETONAT          TABUSM-60-SDI
RMINATION OF URANIUM BY SOLVENT EXTRACTION AS ACETYLACETONAT          TABUSM-60-SDU
                        SOLVENT EXTRACTION OF IRON WITH CHLO          TABUSM-60-SEI
                        SOLVENT EXTRACTION OF METAL ACETYLAC          TABUSM-60-SEM
                        SOLVENT EXTRACTION OF MINERAL ACIDS.          KERTAS-60-SEM
                        SOLVENT EXTRACTION OF SCANDIUM, YTTR          WARRCG-60-SES
                        SOLVENT EXTRACTION OF URANIUM WITH C          TABUSM-60-SEU
TIVE AGENTS IN OILS AND SOLVENT EXTRACTS FROM WOOL.                   HOBSBC-60-MDN
MINERAL ACIDS.  SOLUTE- SOLVENT INTERACTION IN THE SYSTEM HY          KERTAS-60-SEM
 ZEROII TO REACTIONS IN SOLVENT MIXTURES OF LOW POLARITY.             ROCEKJ-60-KOF
              EFFECT OF SOLVENT ON N-PI* ABSORPTION SPECTRA           ITOM  -60-ESN
BORANE-II IODINATION IN SOLVENT.                                      HILLMM-60-CDI
ITRATE IN WATER-DIOXANE SOLVENT.   (ENG.)                             ZIMMHK-60-OEC
              EFFECT OF SOLVENTS ON THE ELECTHONIC SPECTRA O          ZHMYIA-60-ESE
ANHYDROUS POLAR ORGANIC SOLVENTS.                                     JANZGJ-60-CHM
IODINE COMPLEX IN INERT SOLVENTS.                                     DEMANM-60-ICI
LKOXY GROUPS IN VARIOUS SOLVENTS.                                     MOODRB-60-NMR
LYTES IN LOW DIELECTRIC SOLVENTS.                                     RICHEA-60-EIP
N REAL AND HYPOTHETICAL SOLVENTS.                                     MEYERG-60-DLP
 MACROMOLECULES IN GOOD SOLVENTS.                                     HYDEAJ-60-SAA
 SOLUTIONS IN DIFFERENT SOLVENTS.                                     SWINAA-60-CSD
UTIONS IN CERTAIN MIXED SOLVENTS.                                     CHERAS-60-MII
ITRATIONS IN NONAQUEOUS SOLVENTS.  (REVIEW)                           RIDDJA-60-ATN
DINE COMPLEXES IN INERT SOLVENTS.  SPECTROPHOTOMETRIC STUDIE          DEMANM-60-ICI
DINE COMPLEXES IN INERT SOLVENTS.  SPECTROPHOTOMETRIC STUDIE          DEMAPA-60-ICI
L PARTICLES IN STANDING SONIC WAVES.                                  DUKNSS-60-TFA
    OMEGA- OXIDATION OF SORBIC ACID AMIDE.                            WAKABK-60-OOS
IVE EFFECT OF PROLONGED SORBITE ADMINISTRATION IN CHOLINE IN          TSCNLA-60-PEP
BIOLOGICAL OXIDATION OF SORBITE INTO SORBOSE.                         STERNG-60-CIB
IDATION OF SORBITE INTO SORBOSE.                                      STERNG-60-CIB
ULIARITIES OF SELECTIVE SORPTION AND PURIFICATION OF STREPTO          BRESSE-60-PSS
 12- HETEROPOLY-ACIDS.  SORPTION OF CAESIUM ON AMMONIUM PHOS          KRTILJ-60-EPA
RYSTALS.  UNEQUILIBRIUM SORPTION OF WATER ON LEAD IODIDE CRY          MOSKNN-60-WVA
                 ALKALI SORPTION ON WOOD FROM DILUTE AQUEOUS          KLEITN-60-ASW
                        SORPTION STUDIES OF A MODIFIED LOCUS          RUSSVA-60-SSM
           STRUCTURE OF SOTETSU FLAVONE.                              KAWANN-60-SSF
```

FIG. 2.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

(383) PROPERTIES OF ACETYLENIC GROUPS ATTACHED TO THE PYRIDINE NUCLEUS. A. N. Kost, P. B. Terent'ev, T. Zavada, (Moscow St. Univ.) Recd. Sept. 12, 1959. Dokl. Akad. Nauk. S.S.S.R. 130, 326-328(1960).



1)	C8 H7 N	(II)	2-Me-5-ethynylpyridine p. 327
2)	C8 H8 BRN	(V)	2-Me-5-(1-Br-vinyl) pyridine and picrate
3)	C9 H11 NO	(III)	
4)	C14 H22 N2 O	(VII)	R=CH_2-CH_2-N(Et)$_2$ p. 328
5)	C9 H11 NO	(VII)	R=Me, 2-Me-5-(2-MeO-vinyl) pyridine
6)	C12 H16 N2 O	(VIII)	NR_2=4-morpholinyl; 2-Me-5-(2-[4-morpholinyl]vinyl) pyridine
7)	C13 H18 N2	(VIII)	NR_2=1-piperidyl; 2-Me-5-(2-[1-piperidyl]vinyl) pyridine
8)	C13 H18 N2	(VI)	2-Me-5-(1-[1-piperidyl]vinyl) pyridine
9)	C13 H16 N2 O	(IX)	NR_2=4-morpholinyl and picrate
10)	C14 H18 N2	(IX)	NR_2=1-piperidyl

FIG. 3.

three topics: first, the types of organization of correlative indices; second, the coding and logic problems that invariably occur; and third, future possibilities and problems. Throughout I will assume that the characteristics (or description words, etc.) that are to be associated with each item (or document, etc.) have *already been chosen* and I will leave discussion of the problems involved with such choices for other speakers on the program.

• Types of Organization

The most general formulation of the generic and correlative index problem for our purposes can be made in terms of an array of zero and unit elements. (See Fig. 1.) The columns are labeled by the items, and the rows are labeled by the characteristics. A unit in a column indicates that the characteristic in the corresponding row is associated with the item of the column. There are basically two types of manipulative organizations, corresponding to a listing of the columns and of the rows of the array. A well-known example of the former is punched cards, where there is a card for each item and the associated characteristics are represented by punches on the card. (See Fig. 1a.) A well-known example of the latter is the peek-a-boo system where the cards represent characteristics and holes are punched according to the items that are associated with each characteristic. (See Fig. 1b.) The well-known Uniterm system is another example of this method.

By non-manipulative indices we mean coordinate indices which appear in bound book form and require no sorting or rearrangement of cards, looking-through holes, or matching numbers on different pages. Such non-manipulative organizations take advantage of the fact that the number of items whose associated characteristics need to be searched *after* all items with the first characteristic have been found, is relatively small. In such indices a list of items associated with each characteristic is made, and then, for each of these items, the rest of the associated characteristics are listed. (See Fig. 1c.) In the format this arrangement is not always evident, but exists indirectly. Examples are the rotation index, the Scan column index, and the Tabledex index.

Since I will classify these latter indices as "new," let me describe somewhat further the Tabledex index, which may in a sense be considered as typical. In Figure 2 we have listed as an illustration the words associated with five items denoted by number only. We have shown these because they are all to be associated with the word "ocean." The Tabledex index would list these items in a column under the word "ocean" and then list in the respective rows other words associated with the items and that are greater in lexographical order than "ocean." The Tabledex index is used as follows: The several character-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

```
                          items                 manipulative           non-manipulative

                    |
                 ---+------------------
Characteristics     | 0 1 0 0 0 1 0         item       char.          char.
                    | 1 0 0 1 0 0 1
                    | 1 1 0 0 0 0 1         char.      item           item   char. char.
                    | 0 0 1 0 0 1 0         char.      item           item   char.char.char.
                    | 0 0 0 1 0 0 1         char.      item           item   char. char.
                    | 1 0 0 0 1 1 0         char.      item           item   char. char. char.
                    | 0 0 0 0 1 0 1
                                             (a)        (b)                 (c)
```

FIG. 1.

```
4.06   AME GU      Downwind expedition:  ocean     Pacific   US
                                      :
8.05   Camera      instrument   news  :  ocean     photo     Russia     ship       submarine
                                        - - - - - - - - - - - - - - - - - - - - - - - - - -
                                                                                           :
5.07   aurora      cosmic ray Europe     general   Holland   Kon NMI    longitude  meteorol :

                                         ocean     program   satellite  solar      tracking :

5.10   Antarctric biology      CSAGI     dust      general   Int UGG    meeting    meteors  :

                                                             ocean      program    Spain    :
                                                   - - - - - - - - - - - - - - - - - - - - -
4.01   AME GU      compass     gravity   instru-  :ocean     ship       U S
                                         ment     :

Ocean
4.06   Pacific     US

8.05   photo       Russia      ship      submarine

5.07   program     satellite   solar     tracking

5.10   program     Spain

4.01   ship        US
```

FIG. 2.

istic words upon which the coordination is to be made are first put in alphabetical order; then the table associated with the first word is referenced, and the row containing all the other words is associated with the item being looked up. The advantage of the alphabetical ordering is to reduce the size of the rows. (See Fig. 2.) Of course, the way the rows are formed in the first place is by the rotation of the words associated with each item together with a subsequent ordering on first word. The rotation index is therefore the simplest non-manipulative index in this sense. In the Scan column index of John O'Connor he strategically places the characteristics or their abbreviations in columns for easier scanning: the column and abbreviation for each of the characteristics being coordinated is looked up and these columns are scanned for a row that contains all the correct abbreviations. The Tabledex and rotated indices are closed, that is, additions cannot be made easily; on the other hand, the Scan column index is an open-end index.

● Coding and Logic Problems

The problem of the size of correlative indices is always present. Coding of the items and the characteristics can greatly reduce the volume. In addition, the exact method for choosing the code may further help reduce the volume. For example, in the Tabledex index organization, suppose we chose population-based codes for the characteristics where by population-based codes we mean codes in which the code number represents the number of items associated with the characteristic. When utilizing the Tabledex tables, the code numbers are first put in numerical order. The advantage here is that the low-numbered tables, which would contain the longest rows, would have the least number of rows; whereas the large numbered tables, which would contain the greatest number of rows, would have the shortest rows.

Another, better-known technique that can be generally applied concerns overlapping or irredundant coding,

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A	B	C	D	E	F	G	H	I
0	0	1	0	0	0	1	0	1
0	0	0	0	0	1	1	0	1
0	0	1	0	0	1	1	1	1
1	1	0	1	1	0	0	0	1
0	1	0	0	0	0	0	0	1
0	0	0	1	0	0	0	0	1
0	0	0	0	1	0	0	0	1

0	0	1	0	0	1	1	1	1
1	1	0	1	1	0	0	0	1
0	1	1	0	0	0	1	0	1
0	0	0	1	0	1	1	0	1
0	0	0	0	1	0	0	0	1

FIG. 3. Clumping of Characteristics: (Parker-Rhodes and Needham).

first introduced by Mooers. Here let us use marginal punch cards as the concrete illustration of a general method—the object of which is to minimize the number of marginal positions required for the codes. Since the codes for the characteristics will overlap, some false drops may occur during a retrieval. Exact formulations have been worked out to determine the probability that a dropped card will be a false drop, and numerical examples indicate that coordinate retrieval on several words will produce an extraordinarily low probability of a dropped card being a false drop.[1] These exact formulations are as follows:

$$p = \frac{Fn - Eh,n}{\binom{H}{y}^x},$$

$$Fn = \sum_{j=1}^{n} (-1)^{j-1} \left[\binom{H}{y} - n - j \right]^x \binom{n}{j}$$

$$Eh,n = \sum_{j=1}^{n} (-1)^{j-1} \binom{H-j}{y}^x \binom{h}{j}$$

where

- p = probability that a dropped card will be a false drop
- H = number of possible carding positions
- Y = number of units in a characteristic code
- X = number of characteristics associated with an item
- n = number of characteristics being searched for
- h = number of positions in logical sum of codes

[1] See "Digital Computor and Control Engineering," by R. S. Ledley, McGraw-Hill Book Co., Inc., New York, 1960.

Quite a different technique for reducing the volume of an index is concerned with the *grouping* of items and the *clumping* of characteristics. By the *grouping* of items we mean the use of the same coding for more than one item. Thus, a collection of items will be the end result of a search, and the interrogator is then required to scan these to choose those he desires. The *clumping* of terms, on the other hand, is designed to take advantage of the following observation: In Figure 3, the letters represent items, and the zeros and units, the item-characteristic arrays may be reduced. Of course a thesaurus must then be utilized in order to reduce the original retrieval words to the clumped word.

Finally, we come to the problem of more complicated searches. Figure 4 describes a technique that is efficient for searching for r out of n characteristics. Similarly, searches can be made where the original retrieval specifications may have to be relaxed in some manner. These techniques for searching under specified conditions can be important aids in more flexible techniques associated with correlative indices.

• The Future

I will not attempt to predict the future of printed-form generic and correlative indices, but certain directions of development appear most likely to occur. The formation of printed indices "on demand," which will have the specifications required by an individual research worker, to be used only once by him, will become more common. Of course, a computer will have to be used to interpret, search for, organize, and print such "on-demand" correlative indices. A new development in computer technology will make such a process more readily available. The new development is the *associative*, or *contenting*, *memory* which will eliminate the need for pre-ordering of the items and/or characteristics. For example, suppose each memory cell recorded an item and its associated characteristics. The contenting memory can search all memory cells simultaneously for other memory cells containing a particular characteristic. A successful comparison will result in the flipping of a detection spot. (See Fig. 5.) Additional characteristics can be searched for, one at a time, and the record of success or failure for each memory word can be recorded on a shift register, as illustrated in Figure 5. We can even go further than this and retrieve all words whose characteristics satisfy a given Boolean function. This can be done by an additional contenting operation where the shift registers will simultaneously search for the appropriate success pattern. Such a development will in the future significantly assist rapid production of special indices.

In summary, I believe that the use of special purpose correlative indices will become more frequent in the future. Bound book form correlative indices will, I believe, become more popular due to their power and convenience, and many further advances will be made in the coding and logic problems.

Moderator Adams: Thank you very much, Robert.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

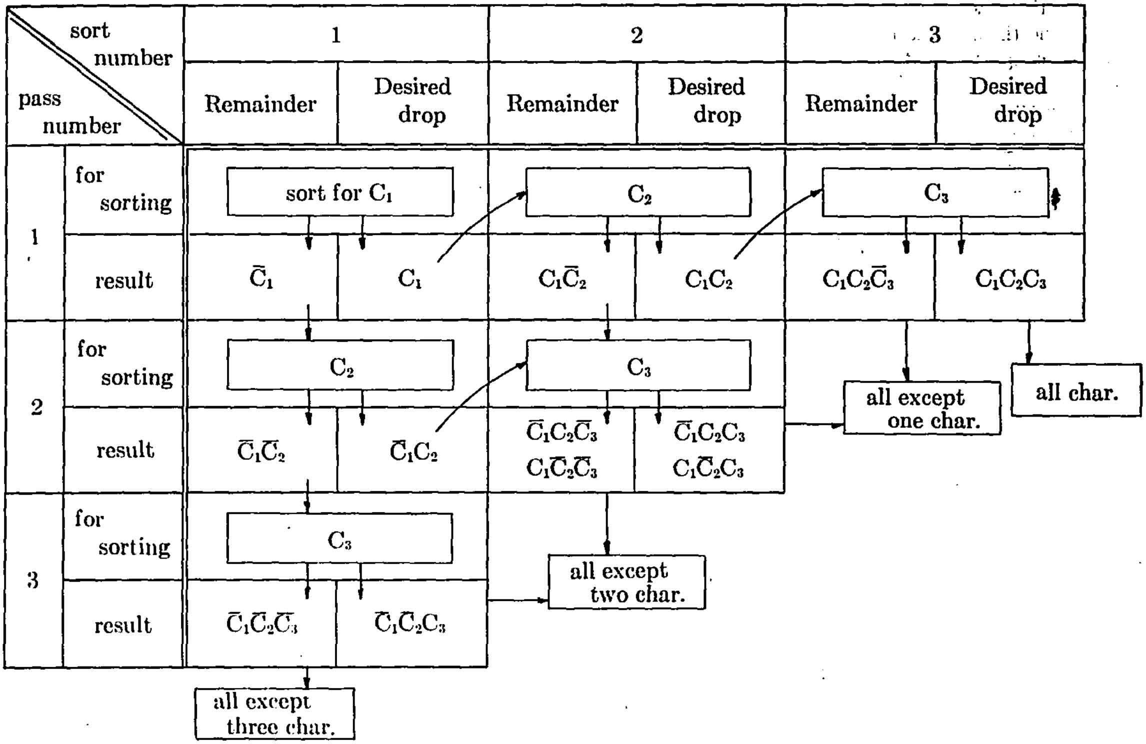


FIG. 4. Searching under conditions.

There are two members of the panel whom I am going to call upon for any voluntary remarks, if they want to make them, bearing on the general topic. Dr. Orr, of the Institute for the Advancement of Medical Communication. Dr. Orr.

Dr. Orr: Our panel originally agreed to appoint three spokesmen to use up the allotted time and I am not going to renege on that agreement.

I just would like to emphasize one thing and that is that as our information retrieval systems become more sophisticated and capable of doing more things, I think we must pay more attention to methods of displaying the services that these sophisticated systems are capable of, and that this is going to be one of the fields in which we will see the largest advances and that people will begin to realize indexes are meant to be used by scientists rather than by a few librarians.

Moderator Adams: Thank you, Dr. Orr. Finally, Mr. Larry Resen, Editor of the *Chemical Engineering Progress*.

Mr. Resen: We at *Chemical Engineering Progress* published by the American Institute of Chemical Engineers, are not in the documentation and indexing business per se, but we are partners in the crime because we are printing the articles in our Journal which have to be indexed. We feel very strongly the extent that John Markus noted that the publishers of the periodicals have a responsibility for providing systematic material for easy indexing.

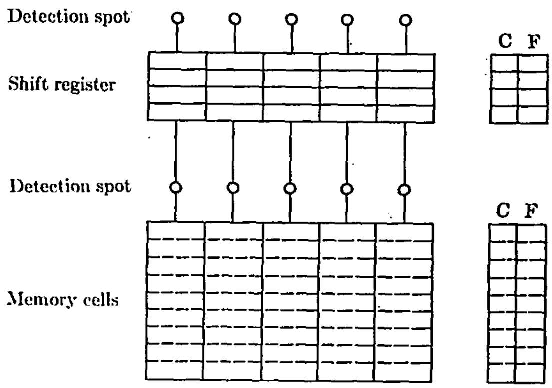


FIG. 5. Associative or contenting memory.

We have adopted the key word concept of indexing. The details on this are available in a reprint for which, coincidentally, the next speaker, Bart Holm, from DuPont, is an author. In another session at this meeting Mr. Rollin Morse, of DuPont, is going to talk about our experience with our system. A part of the system which is very important for the assignment of key words for in-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

dexing, is the thesaurus. One of the most important functions of the thesaurus is to point out and relate generic terms, spelling out what each concept is posted under. We have a well-developed thesaurus available for the use of anyone who wishes to purchase it.

Authors use the thesaurus to assign key words to their own papers. Their key words are edited, along with the rest of their papers, before publication. The key words are then published as part of the article. We feel that our technique will go a long way toward helping to meet the requirement for easily indexed papers.

Moderator Adams: Thank you very much. I want to express my thanks to the members of the panel who not only did not require any moderating in any way, shape or manner, but all of whom were on schedule.[1]

[1] Editor's note: Panelist Malcolm Dyson of Chemical Abstracts Service was absent from the meeting, but he contributed his ideas to the author and other panelists at an earlier time.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.