distribution among 25 of the 31 subclasses between subclass 50 and subclass 80 of patents on the various types of fuel, such as gasoline, diesel fuel, kerosene, jet fuel, etc., and also the distribution among these subclasses of patents on additive functions.

A similar company-class or company-Uniterm correlation will show the areas of activity of any companies selected or which companies have been active in the area represented by a particular subclass or group of related subclasses. The same can be done by company-Uniterm correlations. By company, we mean, of course, the assignee recorded for the patents in the U. S. Patent Office and Uniterm files. These will be available shortly on magnetic tape.

The Indexing of Technical Books*

J. FREDERIC WALKER
Box 27, Haddam, Connecticut

and ROBERT F. SCHIRMER
E. I. du Pont de Nemours and Company, Inc., Information Systems Division, Secretary's Department, Wilmington, Delaware

Received September 7, 1965

The indexing system designed for the ACS monograph "Formaldehyde" (3rd edition) is described. The system, a modified concept coordination, uses a limited number of roles identified by mnemonics.

In spite of recent advances in the technology of retrieving scientific information, the method of indexing chemical texts has remained relatively static. Accordingly, in the preparation of the third edition of the ACS monograph, "Formaldehyde" ("Formaldehyde" III) (*7*), it was decided to employ some of the techniques used in concept-coordination information systems with the object of obtaining a more informative and compact index. At the same time, data-processing methods were employed to simplify and reduce the work involved.

In the following discussion, the index resulting from this work will be described and compared with the conventional index used in the second edition of "Formaldehyde" ("Formaldehyde" II) (*6*). Methods of index preparation including an improved method of assembly developed by Professor L. F. Fieser (*2*) will be reviewed.

"FORMALDEHYDE" III INDEX

A major innovation in this index is the use of a limited number of letters to indicate the function or role of the chemicals, reagents, or materials indexed in specific page references. These function or role indicators (*5*) serve in the place of wordy subentries to designate pages containing information on the material indexed as a product, reactant, catalyst, etc. In addition, small separate indexes were prepared for special chemical reactions, property data on the various forms of formaldehyde, and end use applications.

Before proceeding further, it should be noted that some use of symbols to designate the type of information covered by page references has already appeared in technical books. The use of boldface type to indicate a major page reference is old and well known. The employment of special symbols is noted in Fieser and Fieser's "Organic Chemistry" (3rd ed) (*3*) in which an asterisk is used to denote pages containing physical constants and a dagger sign to designate pages with descriptions of methods of synthesis. Use of multi-part indexes is found in the collective volumes of "Organic Syntheses" (*4*).

The composition of the index of the 660-page text of "Formaldehyde" III is summarized in Table I. The Author Index lists all authors in alphabetical order followed by page numbers in numerical order. Chapter reference numbers, if any, precede the page numbers and are in parentheses. This author–reference number–page number system is standard in many texts.

Table I

Section	Items	Printed pages
Author Index	2955	22
Subject Index		
Explanatory preface		1
(1) Chemicals	1600	15
(2) Reactions	36	0.5
(3) Properties	37	1
(4) Applications	125	1.5

* Presented before the Division of Chemical Literature, 150th National Meeting of the American Chemical Society, Atlantic City, N. J., Sept. 13, 1965.

In the Chemicals section (1) of the Subject Index, seven role or function indicators are used in the form of boldface, mnemonic letters preceding page numbers. A list of the roles chosen is shown in Table II.

The M or miscellaneous role is employed principally for materials of construction, mixture ingredients, and materials cited for comparative purposes. Properties of chemicals other than the various forms of formaldehyde (covered by the Properties section (3)) are listed under P in connection with their preparation or under M.

It should be noted that indexers need not be limited to the roles employed in our work. Roles for physical properties, end uses, patent information, etc., could be added (*5*). Also, some of the roles used by us might not be pertinent to other texts.

Table II

Role	Function
R	Reactant, raw material, input
P	Product of reaction, output
S	Solvent
C	Catalyst, initiator, activator
A	Special agent (*e.g.*, antioxidant, analytical reagent, stabilizer, fungicide, etc.)
I	Impurity, by-product, contaminant, waste
M	Miscellaneous functions not designated by other roles

Comparable sections of our new chemicals index are shown in Table III in contrast with related sections from the conventional index used in "Formaldehyde" II (*6*). Coverage of the two indexes is substantially identical, whereas space requirements are reduced by half with the use of roles in "Formaldehyde" III (*7*).

The Reactions section (2) lists subject matter dealing with name or type reactions, such as the Mannich reaction, the Prins reaction, chloromethylation, etc.

The Properties section (3) summarizes data references relative to the properties of the various forms of formaldehyde, such as formaldehyde monomer, formaldehyde solutions, polymers, and hexamethylenetetramine. Properties, such as density, flash point, free energy, vapor pressure, etc., are listed in alphabetical order with a mnemonic two-letter code indicating the form of formaldehyde involved. Thus, FM is used to identify formaldehyde monomer, FS for formaldehyde solution, PO for polyoxymethylene, TR for the trimer, trioxane, etc. Since this index requires only one printed page, it was thought

Table III. Comparison of Chemical Indexing in "Formaldehyde" III and II

III, Using Roles	II, Conventional Method
Acetaldehyde, I 25, 28, 29, 487, 489-92; M 41, 135, 208, 247, 400, 476-7, 494-5, 497, 624; R 187-8, 222, 249, 285-90, 371, 472, 494-6, 631	Acetaldehyde, 3, 22, 24, 25, 27, 38, 41, 50, 111, 191, 198, 313, 367, 368, 369, 372, 384, 385, 387, 388, 391, 505-6, 512 detection of, in presence of formaldehyde, 373-4 reaction with formaldehyde, 170, 174, 220-3
Aminoacetonitrile, P 241-2, 399; R 242, 295, 399	Aminoacetonitrile, 186 reaction with formaldehyde, 310
Ammonium chloride, A 490; P 526; R 219, 235, 237, 242-3, 365, 385, 423, 251, 453-4, 490, 521, 586	Ammonium chloride, 181, 182, 183, 185, 186, 187, 193, 292, 310, 354, 355, 409, 413, 416, 464
Anthracene, P 437; R 434, 439	Anthracene, 343, 344
Ethylene glycol, M 57, 120; P 244, 634; R 268, 277, 634, 635; S 135	Ethylene glycol, 48, 49, 100, 205, 206, 213, 515
Formic acid, A 154; C 38; I 7, 29, 31, 44, 84-5, 93, 97-8, 106, 126, 490, 497-8; M 498; P 93, 126, 208, 210-2; R 345, 536-7; S 345	Formic acid, 4, 6, 8, 16, 23, 66, 105, 126, 160, 161, 162, 163, 179, 182, 183, 185, 188, 198, 285, 292, 305, 312, 359, 367, 369, 375, 376, 386, 399, 401, 425, 429, 521, 527 effect on polymerization of monomeric formaldehyde, 393 in commercial formaldehyde solutions concentration limits of, 393 determination of, 394 in formaldehyde solutions determination of, 96 formation of, 74, 78
Trioxane, A 587, 608, 624, 636; M 582-3; P 198-9; R 217, 305, 416, 452; S 305	Trioxane, 1, 130, 133, 141, 146, 158, 404 hydrolysis and depolymerization of, 149-153, 366 physical properties of, 34, 116, 146-8 physiological properties of, 153 preparation of, 114, 153 reactions of, 166 180, 192, 195, 205, 212, 236, 273, 286, 326, 355, 356 structure of, 133, 146 sublimation of, 142 thermodynamic properties of, 148-149 uses of, 437, 465, 477, 484, 508

that the reader could readily check the code for the compound form as he made use of it.

The Applications section (4) lists uses and fields of application for formaldehyde in alphabetical order followed by page references.

The index of "Formaldehyde" III does not contain material that can be readily located from the table of contents which is paginated and expanded for this purpose. Such information includes subject matter on formaldehyde production (Chapter 1), handling of solutions (Chapter 3), detection and analysis (Chapters 17 and 18), etc. An example of the utility of this table for anyone wishing to locate information on formaldehyde solution distillation is manifest in the table of contents for Chapter 6 as reproduced below.

Comparison of the two formaldehyde monograph indexes is not appreciably affected by this reliance on the Table of Contents in "Formaldehyde" III, since less than 10% of the 2130 entries in the subject index of "Formaldehyde" II are of the type that would be eliminated in this way. Data on the two indexes illustrating the high degree of compactness attained in our new index are summarized in Table IV. These data show that although the increased size of the text of III is roughly proportional to the increased number of literature references (*viz.* 24 to 30%), the number of words and printed pages is 38 to 39% less. In connection with this comparison, it should be noted that the number of words per page is practically identical for the two editions. Analogous pages on which there are no chemical formulas or section headings contain 420 to 425 words.

Table IV. Comparison of Subject Indexes for "Formaldehyde" III and II

	Edition		Differences
	III	II	% based on II
Pages of text	660	553	+24
Literature references	2947	2273	+30
Subject Index pages	18	29	−38
Subject Index words	3089	5044	−39

Although it is recognized that compactness in an index is not as important as informativeness, and informative compact index is certainly easier to use and less expensive to print than a bulky one.

INDEX PREPARATION

Conventional Method. The conventional procedure for index preparation is usually carried out by means of 3 × 5 cards. The author goes through the page proof and enters each item to be indexed with its page number on a separate card. When this has been done, the cards are placed in alphabetical order following the main entries using a card file to facilitate arranging. Sub-entries are then combined in alphabetical order following each main entry, and a line is drawn through the main entry on the card. When this is done, the cards are submitted to the printer, if legible, or preferably the index is typed out with all sub-entries properly indented.

Fieser Method (2). This method is initiated on receipt of the first batch of galley proof. A copy of this proof, which has been marked with all corrections and additions, is taken by the author who underlines the items to be indexed. Special entries generated in this step are entered on the proof margins, *e.g.*, "Phenanthrene, sulfonation," etc. A secretary then types each underlined or added entry on the top line of a 3 × 5 card and adds an identifying notation (chapter number, galley page, etc.) a few lines below. As these cards accumulate, they are alphabetized, and duplicates, triplicates, etc., are combined on the first card for each entry in a series. When the last galleys have been corrected and indexed, the cards are mounted on legal-size paper by fastening the back of the lower end of each card to the paper with 0.5-in., double-coated, pressure-sensitive tape so that all that shows when the cards are flat is the top of each card with the entry that is to appear in the final index. Two copies of the card paste-ups are then prepared using any standard duplicating equipment. These copies make an index that is complete except for page entries. On one of these copies, the author enters the proper page number as soon as the page proof arrives. For this purpose reference is made to the notations on the cards in the original paste-up which can be folded back for inspection. The secretary then types the page numbers on the other duplicate copy which is sent to the printer.

The author index is prepared at the manuscript stage. A card is made for each author cited with the appropriate chapter or section noted on the lower part of the card. New references to the same author are added as they appear. In addition to its use as the eventual copy for the index, the growing card file makes a valuable reference by means of which the author may review citations as the book progresses.

The Fieser procedure not only simplifies the conventional card indexing technique but also greatly reduces the time required for indexing the book when the page proof arrives. Indexing from the galley proof could also be used with our computerized process (see following section), wherein the keypunch operator would replace the secretary and a preliminary printout could be completed by the author on receiving the page proof.

"Formaldehyde" III Index Preparation. The index of "Formaldehyde" III was prepared from page proof. Proof as marked by the author was used for the preparation of keypunched IBM cards. The cards resulting from this work were then sorted alphabetically and printed to give a machine listing that was checked and corrected by the author before submission to the publisher.

In the case of the author index, use was made of the proof pages containing the alphabetical author reference lists at the end of each chapter. The page numbers for each reference were written on the proof just before the

reference numbers preceding the authors' names. These pages were then keypunched on IBM cards with each author's name, the reference number, and the page numbers. In dealing with a reference involving several authors, a separate card was punched for each author. The finished printout listed individual author names in alphabetical order followed by reference numbers in parentheses and page numbers. Reference to a few authors whose names were cited in the text without specific references were listed separately for the keypunch operator on the upper or lower margins of the proof. Thus, an index item such as "Wagner, K. (79) 45; 113" would indicate that Wagner's work in a specific paper was cited on page 45 with respect to reference number 79 for the chapter involved, and a general citation to his work without a reference occurred on page 113.

In preparing the Chemicals index, the author underlined the chemical or material to be indexed on the page proof and wrote the role letter on the nearest adjacent page margin. The first portion of the name underlined was boxed if it was to be rotated. All numbers and locants, appear as, *e.g.*, "dinitrophenol, 2,4-," and "butanol, iso," in the index. This provided for computer alphabetizing and did not allow locants to determine alphabetical order. Unwanted words were crossed out. Thus, a statement in the text such as "acetate of cellulose" was marked "acetate ~~of~~ cellulose" so that it would be keypunched to appear as "cellulose acetate" in the index. If a chemical was discussed in two or more functional roles on a given page, the several role letters were marked in the adjacent margin separated by commas. The marked page proof was delivered to the keypunch operator for preparation of IBM cards containing the chemical name, role or roles, and page number of each item underlined. "See" references were generated as needed and written on a separate sheet of paper with no page numbers for keypunching and sorting. This is illustrated by index items such as "Methenamine. See Hexamethylenetetramine" and "Methylolphenol, ortho-. See Saligenin."

The small sections of the subject index dealing with reactions, physical properties, and applications were prepared by the conventional card technique.

Since this work was experimental, the keypunched cards for the author and chemical indexes were sorted, processed, and printed using standard tabulating equipment. However, this work is applicable for processing on small-scale computers such as the IBM 1401. It is estimated that IBM 1401 computer time required for preparing an index of 10,000 items or page references (comparable to the "Formaldehyde" III index) from keypunched cards would require 30–40 minutes or, at current computer rates, $25–35. The programming effort to write a computer program to sort, process, and edit such an index is estimated at 3–5 days for an experienced programmer.

The 10,000 items in the "Formaldehyde" III index required about 90 hours for keypunching and verification. Chemicals required almost 50% more time per item for keypunching than authors because of the operators' unfamiliarity with chemical nomenclature.

Although the total keypunching, programming, and computer costs for preparing an index on a one-time basis might be difficult to justify, such costs could be justified by publishers as a service offered to authors. A generalized computer program for index preparation would be a one-time cost. Keypunch and computer costs would be offset by a reduction in manual indexing effort in addition to furnishing the printer with 8½ × 11 in. computer printouts with indentations and punctuation indicated for typesetting. It is anticipated that future advances in optical scanning techniques will replace the necessity for keypunching and will result in a significant reduction in cost.

A computerized indexing process would assure publishers of speedy delivery of clean, uniform index copy and save authors valuable personal and secretarial time.

READER REACTION

In order to obtain some idea as to the utility of the "Formaldehyde" III index from the point of view of the user, opinions on the subject index were solicited from a group consisting principally of industrial chemists who would be most likely to use the book. In this connection the authors are grateful to E. L. Barkley of the Heyden Newport Chemical Corp., Katherine S. Chase of the Hercules Powder Co., R. L. Wakeman of the Onyx Chemical Corp., M. A. Naylor of the Industrial and Biochemicals Dept. of the Du Pont Co., and H. L. Wampner of Reichhold Chemicals Inc., for supplying comments of one to three individuals.

Since reader reactions should be of value to future indexers, representative critical statements, obtained in this sampling, are reproduced below:

(1) "It is necessary for a person who employs the book only occasionally to review the indexing system almost each time he uses it. In such cases, the subdivisions and classifications are of little help and may be a hindrance."

(2) "When I became familiar with the code system, I found the method convenient and rapid. At first, it was rather confusing."

(3) "My personal preference is to have only one index."

(4) "The index ... is more useful than the general type because it is not cluttered up with all sorts of minor references. Most conventional indexes (are) rather difficult to use because of their lack of simplicity and selectivity."

Only one formal book review, of which we are aware, criticized the index specifically. This criticism was as follows (*1*):

"Fortunately, the table of contents has been greatly expanded and paginated. If this had not been done, it would be exhaustive to try to locate data by using the four-part index, the complexity of which is exceeded only by the 'International Critical Tables'."

CONCLUSIONS

Comparison with conventional indexes and a sampling of reader reaction indicates that the use of role or function indicators in indexing technical books results in a more compact and efficient index. Multipart indexes and expanded, paginated tables of contents are of definite value for some texts. Disadvantages in new indexes result from unfamiliarity on the part of the occasional user and should disappear with the increasing general knowledge of information science.

Improved methods for the preparation of an index from marked printers' proof have been developed. These methods are adaptable to the conventional process using 3 × 5 cards or the more sophisticated data processing techniques.

A generalized computer program for index preparation could probably be justified by publishers as a service to authors. This would assure rapid production of uniform index copy and save personal and secretarial time for both authors and editors.

ACKNOWLEDGMENT

The authors wish to express thanks to Professor L. F. Fieser for permission to publish his improved method of index assembly in this paper. Thanks are also due to K. H. Zabriskie of The Chemical Abstracts Service for his help in initiating the preparation of our index, to E. I. du Pont de Nemours & Co. for their cooperation in the project, and to Gessner G. Hawley of the Reinhold Publishing Co. for making it possible to use a new type of index in the third edition of "Formaldehyde". We are also indebted to Reinhold for permission to reproduce portions of the index in this paper.

LITERATURE CITED

(1) *Chem. Eng. News*, **42**, 62 (Oct. 19, 1964).
(2) Fieser, L. F., private communication (Sept. 29, 1964).
(3) Fieser, L. F., Fieser, M., "Organic Chemistry," 3rd ed, Reinhold Publishing Corp., New York, N. Y., 1956.
(4) "Organic Syntheses," Collective Volumes, John Wiley and Sons, Inc., New York, N. Y.
(5) Walker, J. F., *J. Chem. Doc.*, **4**, 45 (1964).
(6) Walker, J. F., "Formaldehyde," 2nd ed, Reinhold Publishing Corp., New York, N. Y., 1953.
(7) Walker, J. F., "Formaldehyde," 3rd ed, Reinhold Publishing Corp., New York, N. Y., 1964.

Processing, Publishing, Storing, Correlating, and Retrieving Biochemical Information at Chemical Abstracts Service*

KENNETH H. ZABRISKIE, JR., and MICHAEL F. LYNCH
Chemical Abstracts Service, The Ohio State University, Columbus, Ohio 43210

Received July 7, 1965

Recent developments at CAS in collecting, publishing, storing, manipulating, and retrieving biochemical information in a computer-based system are reviewed. The relation of this system to other operations at CAS is shown.

I. INTRODUCTION

This paper describes how biochemical information is handled at Chemical Abstracts. In particular, it describes a newly developed operation, the "Chemical-Biological Activities Information System" (called CBAC), and compares it to the way CAS deals with biochemical information in the regular issues and indexes of *Chemical Abstracts* (CA). An explanation of the methods used in CBAC for chemical compound identification and for vocabulary control is followed by a discussion of how CBAC will function both as an alerting service and as a retrospective search program.

II. OBJECTIVES OF THE CBAC INFORMATION SYSTEM

The primary purpose of establishing the CBAC System is to provide rapidly concise and well-indexed information of chemical-biological interest, exclusive of the plant kingdom.** Equal to the primary purpose is the aim of accumulating from the same analysis steps a store of organized chemical-biological activity information, suitable for retrospective searching from a wide variety of viewpoints.

* Presented before the Divisions of Chemical Literature and Medicinal Chemistry, Symposium on Drug Information, 148th National Meeting of the American Chemical Society, Chicago, Ill., Sept. 1, 1965.

** Note Added in Proof. The plant kingdom has been included in CBAC coverage since Jan. 1, 1966.