CHEMICAL INDEXING: THE RESEARCH CHEMIST'S POINT OF VIEW*

By LOUIS SCHMERLING
Research and Development Laboratories Universal Oil Products Company, Des Plaines, Illinois

INTRODUCTION

In discussing indexing from the research chemist's point of view we shall be concerned with the relationship of indexing to the chemist's personal search in comparatively narrow fields. For the purposes of this paper we shall assume that he, like most of us in industry at least, has available a group trained to make a professional search of the literature with regard to broad fields. Also, because I am an organic chemist, I shall here emphasize indexing from the standpoint of organic chemistry; I believe that most of my comments, including even the relationship of nomenclature to indexing, will apply equally well to other fields of chemistry.

The organic chemist will usually turn to an index because he is looking for information on a compound, on a type of reaction, or on a catalyst for causing the compound to undergo a certain reaction. He will usually have in mind questions such as: What are the physical and chemical properties of compound X? Has compound X ever been treated with compound Y or catalyst A? How does one convert compound X to Compound Z? How does one carry out a given conversion (for example, alkylation, dehydration, halogenation, oxidation, epoxidation, nitration, sulfonation, etc.) of a compound similar to type A? What is the mechanism of the reaction of X with Y? Has the thermodynamics of the reaction been studied? One may expect to find answers to these questions, if indeed they have been published, in periodicals (both scientific and trade journals), in technical books, and in the patent literature. While the index of a particular periodical may in itself be adequate and may be turned to if the chemist has an idea of the approximate period of publication, it is, of course, usually to the abstract journals that he turns for an index to the combined periodical literature, and we shall, therefore, begin our discussion with these.

ABSTRACT JOURNALS

The American chemist's most important indexes are, of course, the annual and collective (decennial) indexes of Chemical Abstracts, which is truly the "Key to the World's Chemical Literature." In some cases reference also may be made to Chemisches Zentralblatt, and to British Abstracts, which, however, ceased publication as such in December, 1953.

Chemical Abstracts Author Index.--Depending on the type of information desired as well as on the information already at hand, the chemist may turn first to the Author Index, the Subject Index, or the Formula Index. If he believes that he has seen the desired material in a paper by a particular author published in a particular year or thereabouts, he probably will find the reference most quickly by turning to the author indexes for the years in question.

Use of an author index would seem to be fool-proof. After all, there is no problem of choice of index word or nomenclature involved. The sole problem of both the compiler and user of the index is one of alphabetizing. Nevertheless, a minor difficulty may sometimes arise if the system is not fully understood. For example, I can recall being somewhat perturbed on an occasion when I was looking for a certain publication by Professor Herbert C. Brown of Purdue University. The index name "Brown, Henry," was followed by "Brown, Herman," instead of by "Brown, Herbert C." Since I knew that Herb Brown had published many papers during the year, it was obvious that I was looking for his name in the wrong place. A glance down the column showed the name "Brown, H. C.," and the desired reference. A few moments consideration showed that the indexer had used a logical system which I had not previously had occasion to note. The H. Browns having only one given initial are listed first followed by those having two initials. Indexing is alphabetical by initials rather than first names. Full first or middle names are used, if known, only when necessary to distinguish between authors having the same initials. When the full names are the same, the author's town is included in parentheses after his name. This is probably as sensible a scheme as any for listing names and is quickly understood once called to the user's attention. The system is illustrated in these index names:

Brown, H.	Brown, H. B.
Brown, Harold	Brown, H. C.
Brown, Harrison	Brown, H. D.
Brown, Henry	Brown, J.
(Clitheroe, Engl.)	Brown, James Boyer
Brown, Henry	Brown, James Brooking
(Detroit, Mich.)	Brown, John B.
Brown, Henry	Brown, J. C.
(Madison, Wis.)	Brown, John F.
Brown, Herman	Brown, Joseph F.
Brown, H. A.	Brown, J. G.

*Presented before the Division of Chemical Literature, ACS Meeting, New York, N. Y., September 11, 1957.

The H. C. Brown listing presumably will be changed to Herbert C. Brown in the 1958 index because the name Henry C. Brown also will be indexed.

The Chemical Abstracts Author Index, then, provides a rapid means of obtaining information when one is familiar with the names of the workers in the field. It is, we may add, particularly useful during the interim which must of necessity exist before the appearance of the "Subject Index," as is the semi-monthly author index.

Chemical Abstracts Formula Index.--When seeking information about a particular chemical compound, one may turn to either the Formula Index or the Subject Index. The latter is preferable for well-known compounds which have received much study; indeed, when one looks these up in the Formula Index, he is told to "see" the compound in the Subject Index. On the other hand, the Formula Index is preferred by many chemists for the less common chemicals since its use is independent of a knowledge of the rules of nomenclature and indexing followed by the journal. Furthermore, isomers are grouped together in the formula index whereas they may be widely separated in the subject index in which they are named systematically. For example, in the 1955 Formula Index three chloroheptanes are listed under $C_7H_{15}Cl$: "heptane, chloro," "hexane,2-chloro-2-methyl," and "pentane, chlorodimethyl." In the subject index one would have to look in three places: under "heptane," "hexane," and "pentane." There is, nevertheless, an advantage in looking in the Subject Index. The Formula Index merely lists the pages on which the particular compound (for example, chlorodimethylpentane) is described. The Subject Index, on the other hand, makes mention of the type of information given in each item listed; for example, "2(and 3)-Chloro-2,3-dimethylpentane, formation from tert-BuCl reaction with propene" and "3-Chloro-2,2-dimethylpentane, attempted synthesis of." Such a listing usually will be sufficient to permit the user to conclude whether it is worthwhile for him to look up the abstract.

The Formula Index does have an important advantage in that it lists the C.A. index name of the chemical compounds, thus making it unnecessary for the user to have expert knowledge of the C.A. nomenclature system.

Chemical Abstracts Subject Index.--As is annually indicated in the introduction to the Subject Index, reading of the introduction is not essential for use of the index. Indeed, it is my guess that only a very small proportion of the users have ever bothered to read more than a paragraph or two of the introduction. Perhaps the sole reference made was to determine the meaning of the asterisk or dagger following the names of some compounds. Many undoubtedly have studied the discussion of the nomenclature of chemical compounds which appeared in the Introduction to the 1945 Subject Index, but relatively few have read the discussion of the general policies of the index.

It is axiomatic that if the user understands the problems and methods of the indexer he will be able to find the sought for information more efficiently. The better his understanding of the index, the more likely will he be to turn to the right place at once when seeking data on a particular compound or reaction. If he is searching for information on a number of alkylbenzenes, he will know that he will find the pentylbenzenes under "Benzene, pentyl" and the hexylbenzenes under "Hexane, phenyl." If he is interested in the unsaturated compounds, he will know that allylbenzene is found under "Benzene, allyl" while the butenyl and pentenylbenzenes are found under "Butene, phenyl" and "Pentene, phenyl." On the other hand, even if he is unaware of the principles involved, the ample use of cross references will inform him of the proper index headings. He will then with more or less disgust turn to the proper place, the amount of disgust depending on the particular category into which he falls. The first, and I hope the larger, group are the patient chemists who, when they have guessed wrong, will calmly turn to the indicated location, blaming only themselves for not having a better knowledge of the index. Members of the second group, of whom I know a few, will lose their tempers, will call down the wrath of heaven on the indexer, and will have an urge to toss the index out the window.

The second category also includes chemists, who, without any understanding of or sympathy for the problems of the indexer, feel that the index should be arranged in accordance with their pet personal favorite subjects. They may, for example, feel that all glycols should be indexed under G. They are annoyed by the fact that general references to the compounds are listed under glycols while the individual glycols, except those having common names (e.g., ethylene glycol) are indexed as alkanediols (e.g., propanediol and butanediol). Derivatives in which substitution is on the carbon atom are listed under the individual glycol or diol while those in which substitution is on the oxygen are listed under the hydrocarbon as alkoxy derivatives. These users do not appreciate the value of systematic indexing or of having the individual compounds as major headings rather than as subheadings under glycol.

It is pointed out in the Introduction to the C.A. Subject Index that many factors need to be taken into consideration in deciding on subject entries to be made for an abstract. These include the "author's purpose, his point of view, the new data reported, new or modified methods and apparatus used in obtaining these data, significant relations brought out, theories formulated, new substances prepared, suggested

or likely uses for new substances (as in indexing patents), possibilities for utilization of material (as certain so-called waste material), desirable groupings based on properties, processes or operations, and effects such as industrial poisoning." Because the user's point of view may not always be the same as that of the author and hence of the indexer, the entry will not always be at the point first turned to by the user. Let us look at some examples in the 1954 Subject Index. Assume that we are interested in the reaction of benzene with 1-bromo-3-chloropropane in the presence of aluminum chloride. We will find references under "Benzene, reaction with 1-bromo-3-chloropropane," under "Propane, 1-bromo-3-chloro, reaction with benzene," and under "Aluminum chloride, in reaction of benzene with 1-bromo-3-chloropropane." If we are interested in the reaction of benzene with 1,3-butanediol, we will find a reference, "Benzene, reaction with 1,3-butanediol"; however, if we look under 1,3-butanediol, we will not find the analogous reference to reaction with benzene. Instead, the abstract is indexed under "1,3-Butanediol, hydroxyalkylation with." Obviously, the author's point of view was that of the use of the glycol to introduce the hydroxybutyl group and the paper is so indexed. The subheading or modifying phrase should be such as to let the user know whether it is worth his while to refer to the abstract.

Similarly, let us search the 1955 Subject Index for papers dealing with the alkylation of benzene with propene to obtain cumene. Under the heading, "Benzene," we find "Benzene, isopropylation with propene, increasing yield of butylbenzene homolog-free isopropyl hydrocarbons in." About three columns later, we find "Benzene, reactions of, with propene." Under "Propene" we find "Propene, benzene alkylation with, $BF_3 \cdot H_3PO_4$--$BF_3 \cdot H_2O$ as catalyst in," and "Propene, benzene alkylation with, increasing yield of butylbenzene homolog-free isopropyl hydrocarbons in," and "Propene, benzene alkylation with, thermodynamics of." About one and one-half columns later we find, "Propene, reactions of, with benzene" and "Propene, reactions of, with benzene in the presence of Al and 2-bromopropane." Reference to the abstracts immediately shows that from the point of view of the indexer and of some users, at least, these headings are the proper ones. Nevertheless, if one is interested in every reference to the interaction of benzene with propene regardless of the particular emphasis, he would prefer to have all of them under one modifying phrase. However, much as he would like to see the items of interest to him indexed together, it is rather obvious that this is a Utopian, and usually selfish, goal. As a result, the user must keep in mind the possibility of finding his information in more than one location.

I must admit that from my own selfish viewpoint I probably would usually find optimum use of C.A. if all the abstracts dealing with reactions were indexed merely in the form "Compound A, reaction with compound B in the presence of Catalyst C" rather than having the occasional "Compound A, -------ation of (e.g., alkylation or halogenation, or oxidation of)." But I am really not bothered much by the fact that the references I seek are not bunched together. I believe that one of the important bonuses in using the index (and abstract volumes) is that I often see additional index headings (and abstracts) which, while at the moment are not directly related to the subject at hand, serve to stimulate new ideas. These extra dividends more than make up for the time spent in checking all possible variations of the index words. It is also a good reason why the research chemist should, as often as feasible, do some of his own searching rather than having it all done for him.

It is worthwhile to consider whether C.A. might expand its services by including (hence, indexing) material not at present abstracted. Titles of academic theses, for example, have been recorded, but only when available on microfilm and listed in Dissertation Abstracts. It would be useful to record also titles of theses listed in the American Chemical Society Directory of Graduate Research. Whether the resulting information is worth the added cost in view of the eventual publication of most theses, I leave to the C.A. Editors to determine.

Similarly, C.A. abstracts a review paper such as those appearing in Chemical Reviews by stating that it is a review with x number of references. Analogous reviews appearing as chapters in a technical book also have been listed occasionally. Such listing is certainly to be recommended since this will result in C.A. being a more complete index of the chemical literature.

Chemisches Zentralblatt.--While I realize that "comparisons are odious," I feel that it is worth while, indeed necessary, for our present purposes to compare Chemical Abstracts with Chemisches Zentralblatt. Both Author Indexes use the same system, even with regard to alphabetizing author's names. The Formula Indexes, however, are quite different. Chemical Abstracts uses the Hill system; Chemisches Zentralblatt, the Richter system. Each gives a unique arrangement of empirical molecular formulas and neither has any marked advantage over the other. On the other hand, the type of information given in the two indexes does differ. The Chemical Abstracts Formula Index is fairly independent of the Subject Index and, as already mentioned, merely lists the various isomers by name and gives page references without modifying phrases for all but the very common compounds. On the other hand, the

Formula and Subject Indexes of Chemisches Zentralblatt are not independent. Detailed subheadings are given in either the Subject Index or the Formula Index and there seems to be no way of knowing in advance in which a fairly common compound will appear. For example, in the 1954 Formula Index we find under C_7H_{16} the instructions to see heptane and isoheptane in the Subject Index; six other heptanes are then listed, complete with subheadings and page numbers. Similarly, toluene, xylene and cumene (isopropylbenzene) are indexed in the Subject Index; ethylbenzene, n-propylbenzene, and ethyltoluenes, in the Formula Index. It probably is a safe bet that even the most frequent user of the Zentralblatt cannot guess from time to time where he will find the detailed references for the particular compound of interest. There seems to be little system or consistency. Furthermore, while the Formula Index does have cross references to the Subject Index, the latter contains only a relatively small number of cross references. For example, "Methylamin s. CH_5N" and "Methylbromid s. CH_3Br," but no mention at all of such compounds as ethylbenzene and methylstyrene.

The only hint at the system used is a single sentence appearing on the first page of the index which states, "Those compounds, for which reference to a location in the Subject Index is not made in italics, are found only in the Formula Index." The chemist knows that if he finds a detailed listing of the compound in the Formula Index, he will find none in the Subject Index. On the other hand, if he finds no mention at all of the compound in the Subject Index, he must still refer to the Formula Index in order to determine whether or not the compound is listed.

Unlike the Chemical Abstracts Index, the Chemisches Zentralblatt Formula Index lists only derivatives of carbon. Inorganic compounds are not included. The Richter system was, of course, not intended for inorganic compounds and, unless modified, cannot be applied to them. The Hill system, on the other hand, applies equally well to both organic and inorganic compounds.

A compound's name in the Zentralblatt Formula Index usually is followed by its melting point or boiling point. This feature obviously is a carryover from Richter's "Lexikon." Whether or not such physical data are important in an index is debatable. It does, of course, give the index volume additional use as a reference volume, but one which cannot be relied on definitely since there is no assurance that a particular compound will be listed. Chemical Abstracts gives melting or boiling points only in the case of unnamed (unidentified) organic compounds.

In many cases Zentralblatt lists more than one name for the same compound. Thus, there is the entry "C_2H_5OCl Athylenchlorhydrin (Chlorhydrin, Glykolchlorhydrin, 2-Chloräthanol, β-Chloräthylalkohol)." Chemical Abstracts, on the other hand, gives only one name, the index name, since one of the primary purposes of the Formula Index is aid in the location of compounds in the Subject Index. The Formula Index of C.A. lists the compound just mentioned as "C_2H_5ClO (see also Ethanol, chloro-)."

As indicated above, the Chemical Abstracts Subject Index has the disadvantage for some users that it in many cases fails to group similar things in one location. Such grouping is inherent in the Zentralblatt Formula Index (except for those compounds where the user is referred to the Subject Index). For example, under C_4H_9Cl we find five names indexed: butyl chloride, n-butyl chloride, sec-butyl chloride, isobutyl chloride, and tert-butyl chloride. The subheadings are arranged in order similar to that used by Beilstein: preparation, physical properties, and reactions with various substances. The Chemical Abstracts entry reads "C_4H_9Cl See Butane, chloro-; Propane, chloromethyl-." The four isomeric butyl chlorides are thus divided into two groups, one indexed at "Butane" and the other at "Propane." The modifying phrases are again in alphabetical rather than "Beilstein" order.

The Subject Index of Chemisches Zentralblatt, like the Formula Index, groups or classifies the references. Thus, using again the example of the reaction of benzene with propene we find in the 1954 Subject Index under the heading "Benzol" a group of entries concerning the reaction with olefins as such, with olefins in the presence of metallic aluminum and halogen derivatives, with propylene in the presence of $B_4O_7(H_3P_2O_6)_2$, of catalysts of the system $H_2O{\cdot}BF_3{\cdot}H_3PO_4{\cdot}BF_3$, and of sulfuric acid, and reactions with propene polymers. Under propylene we find references to four abstracts on its reaction with benzene and one to the reaction of propylene polymer with benzene. No modifying phrases as to the type of catalyst in the reactions with propylene are mentioned here.

Those who prefer to have their references grouped together probably prefer the Chemisches Zentralblatt system. There is, nevertheless, an important feature of the C.A. index which I believe outweighs the possible advantage of the Zentralblatt classification system, namely, the fact that C.A. uses a single line for each entry whereas Zentralblatt uses running paragraphs. Chemical Abstracts is much more readily scanned than is the German index and one probably can go through a long column of modifying phrases more quickly than he can read through the solid block of fine print in the particular paragraph of the Zentralblatt to find the desired references.

Furthermore, a classification index may overlook topics that are related in the mind of the user but not that of the indexer. For example, in the Zentralblatt Subject Index subheadings on cycloalkylation and halomethylation are grouped under alkylation as are academic papers on the alkylation of hydrocarbons; on the other hand, the user is told to refer to the section on hydrocarbons for abstracts on technical alkylation and dealkylation of hydrocarbons. For O-alkylation of alcohols and phenols he is told to see ethers; for S-alkylation of mercaptans, he is referred to sulfides. Cyanoalkylation is indexed under addition reactions. Other alkylations are included with Friedel-Crafts reactions. The Chemical Abstracts index heading "Alkylation" includes the statement "(See also Aralkylation, Chloroalkylation, Cycloalkylation, Isopropylation, Methylation)." Neither index mentions, for example, hydroxyalkylation. Obviously, the user must think of all possible index words and use the index accordingly. He should not expect to have his search made for him by the compiler of the index; he should merely hope that he has a reasonably good understanding of the index system and should not be annoyed or lose patience if he has to look in more than one or two places for the information desired.

Chemisches Zentralblatt would be easier to use if it adopted the C.A. scheme of adding a letter to the page number to indicate the part of the page on which the information sought is printed.

Patent Indexes.--The chemical information of patents is, of course, indexed like any other chemical subject matter. The inclusion of the letter "P" with the page number in Chemical Abstracts and of the asterisk in Chemisches Zentralblatt informs the user that the abstract is that of a patent.

The numerical index of patents obviously is of no use when the user does not know the patent number. Conversely, it helps him to get a quick idea of the patent content when he has the number of the patent.

It is fortunate that the abstract journals abstract patents. The United States Patent Gazette has weekly and annual indexes, but these are only author (inventor) indexes and are of little help in searching for patents on a particular subject. Before the middle of 1954, the annual indexes, as well as the weekly indexes, listed the patentees and the title of the patent. Beginning in the midyear of 1954 and continuing through 1955 the index contained only the names of the inventor and his assignee and the patent number. This saved a lot of space; the 1955 Index of Patents Issued from the United States Patent Office has only 449 pages devoted to the list of patentees. However, this manner of indexing apparently met with objections; beginning in 1956 the patent title as well as the names of the inventor and assignee is given. The list of patentees and titles in the 1956 Index occupies 886 pages.

Of course, the patent literature chemist will have available the U.S. Patent Office Manual of Classification, a list of classes and subclasses by general topics, narrowing to the more specific. He will proceed directly to the patents of the desired sub-class.

Current Chemical Papers.--Because British Abstracts ceased publication in December, 1953, I shall say nothing more about it here. It is pertinent to discuss briefly one of its offspring, Current Chemical Papers. This is not an abstract journal but does have some analogous uses. It is an index or, more strictly speaking, a table of contents of current literature, the current signifying literature only one month old. By publishing the titles of papers Current Chemical Papers calls attention to new work more quickly than an abstract journal can. The editors state that it is intended as a guide to chemists' reading and that therefore no index is published. The major advantage of the publication disappears, of course, as soon as the various abstract indexes appear.

The journal divides papers into thirteen classes including the chemical aspects of subatomics, phase structure and phase relation, colloid and macromolecules, reaction kinetics and reaction mechanism, thermodynamics, inorganic, organic, analysis, and apparatus and techniques. The organic section is subdivided into: aliphatic, amino-acids and proteins, carbohydrates, homocyclic, terpenes and steroids, heterocyclic, organometallic and organometalloid, and miscellaneous. The page layout is such that it is surprisingly easy to scan the titles and pick out the papers of interest.

I know some research chemists who prefer to spend their time glancing through Current Chemical Papers rather than through Chemical Abstracts because of the promptness with which the papers appear in the British journal and because of the speed with which it can be "read." Actually, of course, the two supplement each other.

TECHNICAL BOOKS

It seems unnecessary to point out that technical books, particularly those of more than a few hundred pages, should have good indexes if they are to be used efficiently. Depending upon the type of book, the headings in the subject index can be largely the names of compounds or of reaction types or theories. Liberal use of subheadings is highly desirable. Author indexes are also quite useful particularly when the user's interest lies in obtaining a survey of a particular author's viewpoints.

Subject indexes which have no subheadings whatsoever are usually of relatively little value. Even the most patient chemist will be more than a little perturbed to find the heading benzene, for example, followed by twenty or thirty or even one hundred page numbers to each of which he must turn in order to find the one containing the reaction of benzene with propene. Fortunately, such indexes are in the minority, but they are far from rare.

Sometimes a table of contents will show sufficient classification as to make it unnecessary to refer to the index. The user will prefer to leaf through the portion of the chapter in question to pick up not only the directly desired information but also related points that he might not otherwise consider.

LABORATORY NOTEBOOKS

The internal literature of a company such as the technical reports must also be indexed if full value is to be derived from them. Since these technical reports are in the province of the literature chemist I shall say no more about them. However, I shall describe briefly my personal experience both as an indexer and index user with regard to another type of internal literature, namely, laboratory notebooks.

Because I have been concerned with exploratory research, the investigations I have carried out have been quite varied and it would often be quite difficult to locate a particular experiment without some sort of index. The system which I have adopted is very simple. The first two pages of the notebook are left blank for the index. The index itself is based on the names of the reactants and the catalyst, if used. For example, if one of the subjects under investigation is the reaction of benzene with various unsaturated compounds the index line will have the word "Benzene" at the left followed by a series of page numbers above each of which will appear the name or formula of the other reactant as, for example, propene, allyl chloride, dichloropentane, etc. If the reaction was that of benzene with propene in the presence of various catalysts the heading at the beginning of the line would be "Benzene plus propene" and would be followed by the page numbers above each of which would appear a notation concerning the catalyst or the temperature or other variable. The laboratory volumes are sufficiently small, usually 150 pages, that it is unnecessary to attempt any order for the headings other than chronological. Such chronological order is essential because the index entries are usually made each month whereas the notebook may not be filled for several months. A specific experiment can be found readily by scanning the two index pages for the reaction type, and then looking along the particular line for the desired subheading.

Using the system of names of reactants and catalysts also offers a simple and effective means of indexing many years of laboratory experiments. Such an index was begun for me more than ten years ago by one of my assistants. Lined 3 X 5" cards were used. On the first line is written the name of one reactant as the heading. On subsequent lines the names of the compounds undergoing reaction with the first compound are entered together with a number indicating the notebook volume number and the page number; for example, the number 2627-12 indicates that the experiment is described on page 12 of notebook volume number 2627. A second card is prepared in which the heading on the first line is the name of the second reactant, the name of the first reactant is written on a subsequent line together with the volume and page number. This gives a cross reference to the experiment. When a catalyst is used, the name of the catalyst appears together with the name of the second reactant on these cards and a third card is prepared in which the catalyst name is the heading. After some time each card contains a number of references and gives a rapid summary of reactions carried out with the compound or catalyst.

The cards are arranged in alphabetical order by the name of the compound or catalyst on the first line. The nomenclature of these compounds is usually the systematic name unless the common name seems preferable because it is used continually in the laboratory.

The card index has proved to be extremely useful. It gives a quick answer to questions such as, "We studied the reaction of X with Y in the presence of catalyst A; have we ever carried out an experiment on the reaction of X with Z under the same conditions?"

CONCLUSION

In conclusion, I wish to point out that while the research chemist often may not openly express his appreciation for the indexes which are available to him, he is, when he stops to consider the situation, really very grateful to the indexer. He realizes that chemistry is one of the better indexed sciences and that the indexer is constantly striving to make the indexes even better.

I am sure that the art and science of chemical indexing will continue to make beneficial progress.