

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

By NORMAN G. FISHER

Central Research Department, E.I. du Pont de Nemours and Co., Wilmington, Delaware

My assignment, according to the program, is to present "the literature chemist's point of view" about chemical indexing. Let us change the first word in the title so that it reads, "A literature chemist's point of view" -- my own. Not everyone with the label "literature chemist" will subscribe to my opinions. These views stem from an association with chemical literature work for more than 20 years in a large organization emphasizing fundamental research in chemistry. Personal experience includes frequent use of chemical indexes, occasional design and construction of small-scale special-purpose indexes, and recent participation in a group effort to exploit mechanical aids in the set-up of a few indexes of modest size.

At the start of a discourse on problems in indexing, I think it is fitting to recognize how well off we chemists are -- relatively. The accumulated record of knowledge in many other subject disciplines is less well organized and less accessible than is ours. This is a tribute to those who have labored from earlier times to provide the existing instruments we have for hunting among the records.

I turn now for the look at chemical indexing today, as I see it. If we literature chemists have nothing else in common, we certainly do share a vital interest in chemical indexes and the know-how of using them. Indexes are among the key tools in our workshops. Performance of our jobs depends on them. Naturally we are eager to see any new developments that will aid us in these jobs.

The panel chairman posed five questions as a common framework on which we panelists were to build our discussions. I do not believe there is any single set of answers having equal validity for all interested parties. Each instance of use for an index has a certain individuality of purpose. Each index is built in anticipation of serving some recognized purposes. If these turn out to be a poor match with the ad hoc purpose in an individual case of desired use, the user is likely to consider the index a poor one. Stated more simply, perhaps, I mean to say that our purposes in attempting to use indexes are subject to variation through a wide range whereas most indexes do not possess commensurate flexibility of application.

To avoid the difficulties in the phrasing of really precise answers, I shall consider the questions in terms of principles. Question 1 is: what kinds of recorded information warrant chemical indexing?

An index-user whose concern is solely convenience to himself may be tempted to

answer simply "All kinds." Most of us users, however, recognize that indexes are not just wished or willed into being. A price has to be paid. It includes a measure of brains, effort, time, and willingness of somebody to produce the indexes we use. An index user, occasionally turned index maker, appreciates what it takes. I think we will be wise to modify the answer "All kinds."

My choice for first rank in importance as material to be indexed well is the formal record of any new findings of chemical facts. I mean to distinguish factual from conceptual or interpretive information. This choice is founded on the premise that the substance of factual information does not change with time. On the contrary, conceptual or interpretive information, admittedly very important to understanding and integration of facts, has an ephemeral quality. On a long time scale it is subject to change. With respect to value for indexing, I would rank it below factual information. In last place I would put the kinds of information that have only short-term news value. Here I expect vigorous disagreement from some good friends whose primary interests are in the field of chemical economics -- an example being those concerned with market development of new commercial chemicals, or the implications of a commodity price change on an industrial process operation.

I should like to interject a plea to the indexers in connection with the indexing of new factual information. The plea is to avoid basing any selection of items for discard or omission from indexes on personal notions of relative insignificance or unimportance. One frequently hears the plaintive suggestion that our problems with a fast-growing literature would be eased by separation and discard of the "junk" from the more valuable material. We can all sympathize with the motives behind this suggestion. But one person's "junk" may so likely be another's treasure, and what is useless today may turn out to be most valuable in tomorrow's tomorrow.

The unfolding record of new chemistry appears in various physical forms. We have books, pamphlets, journals, monographs, theses, encyclopedias, individual reports, symposia papers, and so on. I do not believe we are warranted in placing any great emphasis on physical form of documents in attempts to formulate guiding principles for selection of material to be indexed. By this I mean it would be unwise, for example, to say that we should carefully index journal articles and not books, or that we should

*Presented at The American Chemical Society Meeting in New York, N. Y., September, 1957.

have a set of selection rules for indexing peculiar to either form of document. There is no reliable correlation between document form and suitability for indexing. In my judgment, the principle set forth earlier, of relating index worthiness to expected permanence of the validity of information, provides more useful guidance. I suggest it be applied in combination with one other important consideration. As an index user, I deem it important that the document indexed have reasonable accessibility. It is frustrating to be referred by an index to a document to which one cannot gain access at all, or not without great difficulty. This consideration is my reason for being dissatisfied with the indexing of university theses if the theses are only available for inspection at some far distant library and sometimes then only after a time-consuming clearance procedure involving approvals by the author and his professor.

The same consideration of accessibility sometimes may be helpful in the choice of whether to put major indexing effort on the form of new information recorded first or on some reprocessed version such as an abstract, or still later in time, an entry in a review monograph or textbook. I feel sure that accessibility of the companion abstracts contributes tremendously to the value of the Chemical Abstracts indexes. It is impossible, of course, to detach the accessibility factor from other important ones such as comprehensive coverage and the convenience of a common language for all the information. At my place of work, the research record indexed is the report prepared by the laboratory chemist, and not the notebooks from which the report is distilled. There is only the original copy of a notebook; there are many copies of the report. Notebook antecedents of reports are identifiable if needed.

Finally, one cannot answer Question 1 as if this question were completely independent from all other questions. The setting in which an index is expected to operate, the purposes it is expected to serve, the users and their needs, are or should be determinants for each index created.

Question 2 is: What kinds of questions should chemical indexes be designed to answer?

Question 3 is: For what kinds of users should chemical indexes be designed?

I choose to consider these questions together because of their close interrelationship. These are possibly the most important two questions among the five. For any particular index, determination of the "right" answers to these questions and the degree of compliance with those answers in the design of the index are prerequisite for utility. I mentioned earlier my belief that it is impossible to develop precise specifications for indexing that will be uniformly applicable for all index situations. It may be useful, however, to comment on real

cases, selected purposely to be representative of appreciably different situations.

The indexing of Chemical Abstracts, I suspect, is the largest chemical index undertaking anywhere, unless the Russians have a larger one. Dr. Bernier will discuss Chemical Abstracts indexing authoritatively, but I wish to consider some of its aspects for convenience in illustrating points bearing on Questions 2 and 3. It is obvious that C. A. indexes are multi-purpose indexes aimed at serving the interests of all kinds of chemists. To the extent that abilities, facilities, and budget permit, I feel sure that the C.A. staff would very much like to have their indexes responsive to the full range of question types that are likely to have appreciable frequency. It is a fair assumption that some people on the C.A. staff have given and do give study to the problems of accommodating the indexes to the spectra of demands made by index users. No doubt they learn much from volunteered criticism. I do not see, however, how there can be any confidence about knowing what users really want in a comprehensive index without now and then making a canvass of a cross-section of subscribers. The results of such a project would certainly be interesting, not only to C.A. but to many others.

Having just made the point about canvassing a circle of index users, it would certainly be presumptuous of me to say that I can itemize all the different sorts of questions that all fellow chemists are likely to pose to indexes. A literature chemist, however, is in a special position to know a little bit about some questions asked by people other than himself. In much of his work he is involved with other peoples' questions, often as essentially an agent for others in getting questions answered, and in additional cases as advisor on literature-consulting problems. Such problems are supposed to be in his special province. On the basis of such qualifications, I may speak about some kinds of questions I know something about from personal experience.

A very large fraction of questions asked, although subject to infinite variation in the phrasing and in respect to locus in the domains of our knowledge, fits roughly into a common pattern. For convenience I call these simple-address questions. The answer sought, if it exists, is usually an individual piece of information or possibly a modest assembly of a few such pieces, and one expects to find the answer under a single appropriate index label -- again, if there is one. This is the everyday stuff that the practicing chemist works with. Questions may pertain to an individual compound or element, properties associated therewith, a particular reaction, apparatus, a procedural method, a specified author's contribution to an identified subject interest, and so on. Except

for troubles associated with sheer magnitude of numbers, these questions are not difficult in principle for index treatment.

A second type of question reflects the human trait of aiding memory and thought through mental association or correlation of items on the basis of something in common. Perhaps I need only to mention a few expressions to illustrate my meaning: such, for example, as "aromatic properties," "negative groups," "beta-elimination," "volumetric methods," "conjugated unsaturation." In the cases of association of concepts already well recognized -- I mean the association is already recognized -- questions involving them may or may not encounter difficulty in indexes. It depends on the specific question, on whether the association includes large or small territory, and on whether the existing literature on the subject is voluminous. But chemists are continually seeking and discovering new associations and relationships. Development and establishment of such new relationships frequently involves a hunt in the literature for comparison of a laboratory finding of the present with a possible ancestor or analog. The kind of question that seeks to validate or elaborate a prospectively new association often is confronted with real difficulty in conventional indexes. Suppose a laboratory colleague comes to us with the alleged discovery of a new reaction of the olefinic bond. He wants to establish whether his reaction is really new and, if it is, he is interested in whether the literature provides any information helpful to guessing the likely breadth of application for his new reaction. Conventional indexes are not set up to yield us readily a complete catalog of reactions known for carbon-to-carbon double bonds. Therefore, right at the beginning we are diverted from a straight-course approach to the desired answer. We may have to be satisfied with an answer of lower stature and modified complexion versus original intentions.

The process of associating concepts may lead to questions of a higher order of complexity. Sometimes wanted information is specified rather precisely in terms of a confluence or combination of several or many characteristic attributes. If we were able to use a plurality of the attributes simultaneously in the searching process, the desired target would be pinpointed accurately and the search should be efficient. It is exactly this kind of question, plus the need for flexibility in the formulation of the usable combinations, that has enhanced interest in the possibilities of some newer indexing techniques and devices. But that is the territory of later questions.

Question 4 is: What kinds of indexes are best suited, respectively, to the various types of chemical literature to be indexed?

The functions of chemical indexes may be likened to those of guide-maps. A road map or

street map is very useful to help one locate the position of individual highways. But suppose one is a highway landscape engineer who wishes to examine a sampling of highways where oak trees have been planted as part of the beautification treatment. The conventional maps you and I obtain from service stations are of little help to the engineer for his oak-tree question. The purpose of this simple example is to re-emphasize the importance of a reasonable match between the capabilities of an index and its expected uses. It is quite appropriate that we should have different kinds of index tool for different kinds of problem. I shall now try to examine a few familiar types of index with attention to their capabilities and limitations.

The conventional formula index has many virtues. As a device for aiding quick and convenient access to information on individual chemical elements and compounds, it is perhaps unequalled. The apparatus can be simple, the conventions of use are easily mastered, size and growth are not severe restrictions, the main structure is free from the complexities and vicissitudes of nomenclature, and there is no need for continual redesign to keep one up-to-date. In short, the formula index is a nice, neat, sensible tool.

But please note what I said it is good for-- locating information about elements and compounds via one approach only, the kind and number of elements involved. At the cost of effort, time and ingenuity, one can work out ways to answer a limited range of generic or coordination-type questions by appropriate manipulation of some letter and number combinations. But the formula index is a dull tool indeed, essentially useless, for answering many types of question that chemists ask. It is even ineffective for handling some chemical compositions that come close to our concept of a compound. The field of polymers is an outstanding example. Another is non-stoichiometric crystalline "compounds" which are currently in prominence from the rising tide of solid-state studies. Here interest lies in component variations in the order of parts per billion or even lower.

I turn to Chemical Abstracts again for an example of the second type of index -- the dictionary style of subject indexing. This is at present probably the most widely used kind of a multi-purpose index. Dr. Bernier undoubtedly will tell about its good points, in expectation of which, time may be saved by jumping directly to some of the shortcomings that seem to me important.

The alphabetical index, as exemplified by Chemical Abstracts, is simple in principle but becomes very complicated in such a large-scale application. It also requires large abilities of the user for maximum effective use. He must be a good speller. Of course he ought to

be anyway. He must have or must acquire a really imposing vocabulary. He needs to be a near-expert in chemical nomenclature. Chemical nomenclature has become hard pressed to deal adequately with the increasing complexity of known features of chemical structure. The result of much labor with nomenclature problems is an ever-growing labyrinth of carefully defined detail rules and conventions that become increasingly difficult for any single person to master -- at least in his spare time. Is it not fair to ask what is the future of any indexing system inextricably dependent on nomenclature? I cite C.A. as an example only because it is widely known. The limitations of which I speak are inherent, to a varying degree, in the dictionary type of chemical subject index in general.

Nomenclature figures prominently in another inherent limitation of the alphabetical index. An alphabetical index is an ordered linear array of labels under which individual pieces of information are dispersed according to their correspondence with the labels. This operation results in scattering the component parts of an original combination in a source document. The combination itself may have been of meaningful interest in its own right. Each one of the dispersed pieces is a key to the combination, but any one key may be a very popular one that fits into other combinations too. If a searcher attempts to make a better approach to his wanted combination by using two or more of the keys together, by looking for a coordination of entries under two or more of the index labels, he may find that he faces a formidable task. Too often it is practically impossible or prohibitively time-consuming. A similar searching difficulty is encountered when the objective is a bit different. The searcher may want to find information on the basis of a certain criterion which is subordinate to a plurality of the index labels. An illustration may be chosen that simultaneously shows how nomenclature complications can operate in this type of situation. Inherent in our nomenclature system is the requirement of subordinating certain structural features to others in name building. Witness the Chemical Abstracts order-of-precedence of chief functions, not to mention all the other functions presumably of lower grade. If the criteria of a desired search emphasize one of the low-order functions, or still more likely a combination of them, the prospective material may be indexed under labels all the way from "A" to "Z." Incidentally it has become my practice to call this an A-to-Z search situation. When we at my location receive a request that appears to call for one, our normal response is to advise the requester that he attempt to modify the request and settle for something less or different that is still useful.

Punched cards have achieved a position of demonstrated utility in connection with chemical

indexing during the last decade, approximately. Possibly the optimum limits of applicability for the punched card have not yet been defined exactly, but at least some of the uses have been fairly well established. A newcomer to the subject with intentions of using punched cards is well advised to investigate the substantial literature now available for guidance.

In my view, the significant contributions of the punched card are two: (a) providing a tool that expands our ability to maneuver and manipulate index approaches to subject matter, (b) introducing the factor of mechanical assistance in the physical operation of consulting the index. Results from the latter may be realized in the form of beneficial relief from pure toil and/or increased speed of getting answers from an index.

Our experience with manually sortable edge-notched cards has shown them to be useful in two general types of situation. First, in making literature searches that involve assembly of information in a subject field as opposed to collecting all that is known about a single issue, we have found it helpful to assemble the extracted information on punched card forms, usually not more than one reference per card. The punching feature is used as a help to analysis, organization, and digest of the assembled material. The indexing is kept as simple as possible and is custom-designed for each particular case. More often than not, this kind of application represents a one-time use of the punched cards, that is, their function is only to aid work-up of the material. Our literature search products are almost always issued in multi-copy report form. Otherwise, the set of punched cards might have more permanent value.

The second general type of use for edge-notched cards includes a considerable range of what I shall call special purpose indexes of moderate size. Here we wish to exploit the ability of punched cards to provide a multiplicity of index approaches to subject matter that can be used in any combination simultaneously. Such indexes are made with the expectation of continued use over a considerable period. At higher levels, the size of the needed vocabulary and the number of cards in a collection become important limitations on the convenient usefulness of edge-notched cards for this application. The exact numerical values for the limits on both factors vary from case to case. From second-hand information, I understand that many chemists use the edge-notched cards extensively for personal collections of information in great variety.

My experience with machine-manipulated cards is more limited. We are using them for chemical indexing of laboratory research reports, as reported at the Miami meeting. Transposed to a larger scale of magnitudes,

my remarks about manually sortable punched cards are applicable to machine cards, I believe.

It is perhaps worth noting that resort to punched cards as an indexing tool demands a high order of straight thinking in index design and construction. Machine operation is a manifestation of applied mathematics. The governing program that controls the pattern of moves made by the machine must be drafted with mathematical precision if satisfactory results are to be realized. Machines themselves cannot distinguish built-in errors of an indexing system such as can be detected easily by visual inspection of an alphabetical index.

Question 5 is: What new indexing techniques are being developed and what are their potentialities?

I wish I knew the answers to this two-part question, and especially the answer to the second part. My position thus far has been that of an extremely interested reader, and observer, anxious to keep informed about newer developments and eager to investigate new possibilities that look outstandingly good. At the moment a rather bewildering array of gadgetry, newly available, gives exciting promise.