

New Information Services—A Practical Approach*

By W. A. WILKINSON and W. H. WALDO

Information Center, and Agricultural Chemicals Division, Monsanto Chemical Company,

St. Louis 66, Mo.

Received February 21, 1961

In the daily business of performing almost any phase of a technical library operation, most of us have wished that someone would offer this or that service. If so, certainly it would attain immediate success! Usually it's something perfectly capable of being done with available techniques but a little too big or too expensive for a single library or company to undertake. These are simple problems which we face every day. We solve them as best we can individually and go on to concern ourselves with more "impressive" or more "important" obstacles—like structure searching, machine translation, auto-abstracting, etc.

Let's stop for a minute and see what are the possibilities of doing something about a few of these simple things. These few ideas are submitted to stimulate your thinking. Perhaps there is a market here which would induce someone to make some badly needed services available to us all? If appropriate services are offered and they are economical (not necessarily cheap), well-packaged, prompt and accurate, they could be begun at once, and would succeed, I'm sure. Note the success of the *Uniterm Index to U. S. Chemical Patents*, *Index Chemicus* and *Chemical Abstracts*, none of which is inexpensive!

Prompt Photocopies of Any References in the Chemical Literature.—For many of us the search is over when we have at last located a long sought-for entry in a *Chemical Abstracts* index. Sure enough, the abstract indicates that the paper contains just the analytical method that we need. Too bad the library doesn't have this particular Japanese journal, but the librarian will find it—he always does! Several letters and a few weeks later a copy of the article is at last received. Can't something be done to make it a little easier and much faster to get? Couldn't someone assume the responsibility to provide photocopies of any article which has been abstracted in *C.A.*?

Some steps have been taken which have made the situation a little better. For instance, the National Science Foundation has provided funds to ensure that certain foreign chemical journals not received in the U. S. (but abstracted by *C.A.*) would be obtained and made available at the Midwest Interlibrary Center. Recently it was announced that Chemical Abstracts Service would sell photocopies of current, hard-to-get Russian chemical papers, at \$1.00 per page. Eugene Garfield's Institute for Scientific Information has its OAST service at \$1/article, plus 10 cents/page.

These are welcome advances, but what we really need is a sort of clearing house to which we could send orders for copies of any articles which have appeared in the chemical literature. The clearing house would be able

to provide copies immediately from its own holdings or forward the order to a source which they *know* can definitely provide it immediately. It's wasteful and unnecessary for each of us to try here, there, and everywhere to find a few obscure (but important) articles, when one agency could do it more quickly and economically.

Deep Index to Unclassified Government Chemical Research Reports.—It is quite a few years now since the P.B. reports made their first appearance. This was followed soon by *C.A.*'s decision that these reports would not be abstracted or indexed by them, since, theoretically at least, this information would eventually be published, and would then be covered by *Chemical Abstracts*. Later, *C.A.* began to abstract and index selected P.B. reports of a chemical or chemical engineering nature.

Unfortunately, experience and several surveys have shown that many of the reports are never published in journals. Those which are published often appear a long time after the corresponding report issued. Therefore, it has become more important that we are able to locate information when it first appears in the report literature.

Announcement of the reports is quite adequate, but retrospective searching is very difficult. Over the years the Office of Technical Services has provided yearly or semi-yearly indexes. However, these have varied widely in the headings used, and depth of indexing. Unless a chemical compound is the major subject of a report, it is almost impossible to find information about it *via* the subject indexes which are provided. Occasionally subjects can be explored by using related indexes, such as the recently issued ASTIA subject index, or some of the many lists and bibliographies which O.T.S. or other agencies have issued. Some private organizations such as Technical Information Service, Washington, have issued detailed subject indexes in selected fields but no really comprehensive, deep indexes are available. Maybe we need a publication like the *Uniterm Index to U. S. Chemical Patents*, which would cover the declassified report literature, providing prompt, deep indexing? Now that more and more of the U. S. research expenditure is through government contracts, an increasing amount of our chemical research will be revealed first in government reports. Our neglect of these reports could be very costly to our employers and our industry.

Prompt, Deep Index to Foreign Chemical Patents.—A related indexing problem exists with the foreign chemical patents. With some exceptions most of them are indexed eventually by *C.A.* and some of the other indexing services. However, this advantage is offset by an urgent need for very prompt yet detailed indexes. This would reduce duplicate filing of applications, duplicate research and help to improve the position of U. S. chemical com-

* Presented before the Division of Chemical Literature, American Chemical Society Meeting, St. Louis, March, 1961.

panies in world competition. Does the success of the *Uniterm Index to U. S. Chemical Patents* suggest that a similar service is needed for the chemical patents from some foreign patent offices, such as Belgium, Germany, France, Great Britain, and Japan?

Some help might be provided by the proposed ACS publication, *Chemical Patents*.¹ It would provide prompt indexes but, of course, not enough depth of indexing for reliable searching. The lack of coverage of Belgian patents is an unfortunate weakness in *Chemical Patents*. The Derwent Patent Service provides good announcements and translations. Many other announcement, abstracting and translating services are available but here again the indexing phase has been neglected.

Index for Manufacturers' Technical Bulletins.—Many of us, or our co-workers, are cooperating with Chemical Abstracts Service in the preparation of the new *Lexicon of Non-systematic and Trade Names for Organic Compounds*, co-sponsored by the American Chemical Society and the Synthetic Organic Chemical Manufacturers Association. This is a significant step toward the solution of problems associated with manufacturers' technical bulletins and similar trade literature. However, there still will exist the problem of finding specific information (such as physical properties) in individual technical bulletins.

So often when you need to find a certain physical constant for what might be a fairly common compound, much to your surprise (and disappointment) you'll discover that almost all the other constants are found in the handbooks in no time at all. But the constant that you need seems to have escaped the compilers. So, as a last resort before making a systematic search through *C. A.*, etc., you tackle the trade literature collection. Using buyers' guides, dictionaries, directories, and similar compilations, you locate possible suppliers, hunt through their bulletins and if you have a large, up-to-date collection, surprisingly often you will find the property you wanted! Of course, some libraries have indexed their technical bulletins in depth. However, we feel this expense of manpower is too great. For a comprehensive collection, detailed home-made indexes are too expensive when the size of the indexing job is weighed against the probable return, especially when each library must duplicate the indexing efforts of all the others.

Perhaps someone could make a commercial success by beginning a service somewhat like this. Copies of technical bulletins issued by all U. S. chemical companies would be obtained on a regular basis. Each would be identified and the compounds and properties indexed. Subscribers would receive cumulative indexes to all bulletins and sufficient bibliographic information about the bulletins to enable subscribers to obtain them from the companies which issued them. (This would improve both the acquisition and searching of the bulletins.) An added feature might be the sale of microfilm copies of all bulletins indexed, perhaps in a manner similar to the Vendor Specs Microfilm Film system. Such a service might even be subsidized by the chemical manufacturers who issue the bulletins.²

Abstract Copy Service.—Perhaps the most tedious part of any large literature search is the step which follows the selection of column numbers from the *C. A.* indexes. In some cases hundreds of index entries must be consulted. For each column number the searcher must remove a volume of *C. A.* from the shelf, turn to the abstract, read

and evaluate the abstract and, if the abstract is pertinent, all or part of the information there must be recorded. Much of the clerical part of this sequence could be eliminated if inexpensive copies of the abstracts could be obtained quickly by supplying the lists of column numbers to a service bureau, presumably at Chemical Abstracts Service. In his review of the research program at *C. A.*, G. Malcolm Dyson, Research Director, suggested that they hoped to be able to supply services of this nature "at suitable rates."

As long as so-called "manual indexes" are used, and this may be for a long time yet, we must accept their advantages and try to minimize the disadvantages. The elimination of this time-consuming "look-up" step would allow the searcher to concentrate his efforts on non-clerical activities. For instance, another search would be begun while someone else is preparing and sending copies of abstracts already identified. At the same time Chemical Abstracts Service would be able to collect a small copyright fee for each abstract provided. But the copies would have to be of satisfactory quality, reasonably priced, and delivered promptly, maybe even *via* wire service facsimiles.

Display and Print-out Equipment for *C. A.* on Microfilm.—For large volume searching another system would be even better. Automatic equipment such as the FMA Inc. File Search System could be enlarged and perfected to provide rapid display of any abstract (from microfilm) with the option to make quick paper enlargements when needed.

Such a system would require that a complete set of *Chemical Abstracts* be made available on microfilm. This series of reels would be permanently loaded into the search unit. After a column number is punched into a keyboard, the unit would select the proper reel, turn to the page containing that column number, and display the abstract. The literature searcher would read the abstract, reject it by punching the next column number or make a paper enlargement for further study, before proceeding to the next abstract.

Equipment such as this could be used to provide photocopies of abstracts on a service basis as previously discussed. The document retrieval portion of IBM's Project Walnut system could, of course, provide photocopies of abstracts stored on strips of film, if it were necessary to use extremely fast, high capacity equipment.

Single Deposit Collection of Translations from World-Wide Sources.—Previously we discussed the problems associated with obtaining copies of articles once you had references to them. This difficulty is compounded further when a translation of the article is needed. Fortunately (or unfortunately) there are many sources of available translations which should be checked before a translation is begun. The ins and outs of this growing bibliographic maze will be discussed in another paper to be presented to you later. Its author has concluded, as we have, that it should be possible to order a copy of any available translation from one source, to receive it promptly and at a reasonable price. The SLA Translation Center and O. T. S. have made good progress but there remain overlaps with other groups (such as Central Intelligence Agency), and although many foreign sources are publicized, the translations themselves are not routinely imported for re-sale.

Physical Properties Tables and Indexes Service.—The last service to be discussed, and of course there are others

which could be considered, is related to the documentation of physical and chemical properties of materials. Our principal interest is in the properties of chemicals under a wide variety of conditions.

The "International Critical Tables," our "old-stand-by" will not be revised. Instead, the Office of Critical Tables (of the National Academy of Sciences) will coordinate decentralized collections of critical data. For the user, the problem is first to become aware of all sources of potentially useful data, then to obtain publications which might be useful in a given field, and then to be able to use the various systems in which the data are presented in order to locate the value you need.

One group has decided to do something about this. It is the Thermophysical Properties Research Center at Purdue University. Picking up in time where the "International Critical Tables" left off, they intend to collect, evaluate and disseminate all recorded information on the thermophysical properties of all substances. Incidentally, in the research phase of their work they also intend to measure new properties and incorporate them into the literature. Since its beginning in January, 1957, the Center has proceeded to do that which some said couldn't be done. Their financial support has come from industry, which would indicate that some companies are willing to pay large sums for necessary information about properties of materials. More recently, Chemical Abstracts Service has begun plans for a physical properties data service.

The chemical industry has supported, directly or indirectly, several research projects which have generated a large amount of data. The *Journal of Chemical and Engineering Data* was developed by the ACS to publish

in a single place papers that formerly went into their other journals. What we need now is better control and dissemination of properties scattered in many primary publications, possibly by beginning some kind of a specialized secondary publication. Perhaps with industry-wide support a physical properties serial publication could be established. It would issue simple tables of new data collected from primary journals, manufacturers' technical bulletins, research center compilations, etc., organized in serial number order. A separate coordinate index would enable users to relate compound, conditions and property and turn to the correct table for the value. Maybe this wouldn't be such an impossible job?

CONCLUSION

Most of you will agree that there is a need for these services, and probably you could add several to the list. There is no lack of equipment or know-how. All of these services could be performed today accurately, promptly and in an acceptable format. The remaining questions are of economics and initiative. The ACS and other organizations have shown some initiative, yet there is still plenty of room for the private profit seeker. If the vacuum is left too long, however, the Federal Government may have to fill this gap. Perhaps someone else is willing to sell these services and do it now!

- (1) Plans to publish *Chemical Patents* have been abandoned.
- (2) This service has now been developed as the Chemical Specifications Microfilm File, an affiliate of the *Oil and Gas Journal*. The first file and indexes will be issued about August, 1962.

An Algorithm for Translating Chemical Names to Molecular Formulas*

By EUGENE GARFIELD

Institute for Scientific Information, 33 South Seventeenth St., Philadelphia 3, Penna.,
Received March 13, 1962

To calculate a molecular formula, a human or a machine computer must first be able to recognize the chemical name or the structural diagram on which the molecular formula calculation must be based. Prior to the publication of my book¹ on this subject, there has never been a serious consideration of the possibility of computing molecular formulas directly from chemical names. Chemists have always assumed that it is first necessary to draw a structural diagram before the molecular formula of a chemical can be calculated. Furthermore, the vagaries of chemical nomenclature have created the psychological climate that this step must be necessary. It has been axiomatic that in order to obtain the same molecular formula the chemist must work from the same structural diagram. Naturally, when you give it a second thought,

you know this is not true. For example, if I say butane, the average chemist knows its formula to be C_4H_{10} . It is not necessary for him to draw the two dimensional ideograph $CH_3CH_2CH_2CH_3$ or the linear notation to arrive at the correct molecular formula. Once you accept the idea that the structural diagram is not necessary, then you can proceed to the question of how one "recognizes" a chemical name.

The chemist reads a chemical name and has a built-in mental dictionary that tells him certain combinations of letters have a particular referential meaning. For example, butane is a string of four carbon atoms. However, a computer is a far less sophisticated "reader" and must be instructed in a very precise fashion how to "recognize" the occurrence of meaningful strings of letters. However, as chemical names get more complicated, the chemist also has difficulty in identifying meaningful segments of chemical names. It is, therefore, important and very

* Presented before the American Chemical Society, Division of Chemical Literature, March 22, 1962, Washington, D. C.