Training Chemists in the Use of Chemical Abstracts' Services*

RUSSELL J. ROWLETT, Jr.
Chemical Abstracts Service, Ohio State University, Columbus, Ohio 43210

Received December 27, 1968

Teaching the chemist to use *Chemical Abstracts* (CA) and its indexes has come about over the years through a variety of approaches which include papers and publications, encouragement of and participation in college-level scientific literature courses, and the interest of dedicated chemistry professors who lend their experience in guiding the student through use of the literature. A more intensive and formalized type of training is being developed and applied to teach the effective use of the computer services evolving at Chemical Abstracts Service. Traveling seminars, open forums at national meetings, intern programs, and extensive publications on the nature and application of the new systems are discussed as means of instructing the chemist in the modern concepts of chemical information handling.

Almost 20 years ago, a former editor of *Chemical Abstracts* (*CA*), Dr. E. J. Crane, wrote, "An editor needs ears more than he needs a pen when the use of his journal is discussed."[1] I am violating Dr. Crane's directive; however, the rapid and continuing changes in Chemical Abstracts Service (CAS)—and in scientific literature generally—need explanation. The new classes of chemistry students and established chemistry graduates both need guidance in how to use the services which modern technology is providing in the field of scientific information storage and retrieval. This paper briefly reviews some of the recognized CAS attempts to train chemists and looks at some other activities which are not considered normally to be training tools. These latter activities are important in training chemists in the use of the new CAS computer services.

Throughout most of its history, CAS has performed its mission by publishing a single abstract journal. This journal, which is still a printed publication, has progressed from a monthly to a weekly, and includes issue, volume, and collective indexes arranged by subject, author, formula, patent number, and ring structure. Over these years, books by Crane and Patterson[2] and many journal articles by Crane[3, 4, 5, 6, 7] and members of the *CA* staff served to alert the scientific community to the development of the printed abstract journal and its indexes. In more recent years, contemporary authors such as Bottle,[8] Burman,[9] and Mellon[10, 11] have prepared comprehensive studies of the use of chemical literature. Each includes explanatory chapters on the contents of *CA*.

The ACS Divisions of Chemical Education and Chemical Literature have provided platforms for discussions of the growing problems of chemical literature, and these have served very useful instructional purposes. Papers presented during such programs were collected in 1961 as an *Advances in Chemistry* volume,[12] and many make direct references to the content of *CA*.

*Presented before the Symposium on Training Chemists in the Use of the Chemical Literature, Divisions of Chemical Education and Chemical Literature, 156th Meeting, ACS, Atlantic City, N. J., September 10, 1968.

Thus, the printed page, in book and journal format, has been used extensively as a training tool to describe the content of *CA*. Unfortunately, few of these publications detail how to use *CA*. Someone has declared that the use of *CA* is a "diverse and personal process." Perhaps over the years we have been better equipped to list what the abstracts and indexes contain, and to outline the philosophy of how they are put together, than to really teach how to use them.

Throughout these same years, CAS and the American Chemical Society have encouraged colleges and universities to include in their curricula a course in scientific literature. With a few exceptions, this effort has not been eminently successful. One of the exceptions is the course at The Ohio State University in Columbus, where *CA* staff members offer the course as adjunct professors. Several times ACS has considered offering financial incentives on subscriptions to colleges that present a course on how to use *CA*. Recently it was suggested that the use of the chemical literature might be a subject for one of the popular ACS short courses. Such a half-day short course was presented in December 1966 at Indiana Central College, at the request of the college, by a CAS managing editor.

CAS has sought a partial solution to the teaching dilemma by making staff members available for talks to local ACS Sections, one-day seminars, and college lectures. This is a feeble effort when compared to the total needs, but it is all an economy-minded budget has allowed.

There have been dedicated chemistry professors who have organized literature courses and have sold their administrations on offering such courses for different lengths of time. Many of these men have been *CA* abstractors or section editors and, thus, were familiar with the production methods of an abstract journal. In retrospect, these few professors have been the only front-line teachers training chemists in how to use *CA*.

All such teaching courses have the same basic problem: there are very few chemists suitably trained to offer such a course, either for a semester or for a concentrated two

days. Without such courses, how can teachers be trained? It is a vicious circle from which CAS has not been able to escape successfully. Today's science curriculum demands a thorough course in handling the scientific literature. The advent of the digital computer and its use in information processing removes such a course from the elective category and makes it a degree requirement.

It has been assumed that experts in the production of *CA*—the staff of abstractors, indexers, and editors—are also experts in the use of *CA*. In the Editor's best hindsight, this statement is not always true.

Proficiency in searching a large printed index, such as *CA*, comes only through much practice, plus a generous pinch of horse sense. Someone has likened this sixth sense to the instinct developed by a veteran duplicate bridge player to locate the cards which he cannot see. Both are obtained only after many hours of actual experience, and the experience of producing the abstracts or indexes is different from the experience of using them.

There can be no one set of rules for searching *CA*. Where to search depends largely on the specific bit of data being sought and on the detail of this data which is required. Bottle[13] in "The Use of the Chemical Literature" declared, "There are so many purposes for which the *CA* Index can be used that only a series of suggestions for general guidance can be proffered." The philosophy employed by an editor in the preparation of an index is too often different from the philosophy of a researcher using that same index. *CA* editors try to place a subject-index entry at the location where they believe the searcher will look first. A comprehensive list of over 84,500 cross references is inserted to direct the searcher who may have looked first at another location. In a sense, these cross references are teaching tools on the use of the subject index, and CAS is considering publishing them in a separate book for such use.

For 59 years (65 volumes) *CA* Subject Index introductions tried to teach chemical nomenclature, plus descriptions of how the indexing was done. These introductions were textbooks on how the indexes were built. Beginning with Vol. **66** (January 1967), the Subject Index introduction was reoriented to attempt to tell the searcher where to expect to find data. This change was designed to make the introduction a textbook on how to use the index. It is an effort toward better teaching.

Teaching the use of systematic chemical nomenclature has always been an objective of CAS, carried out primarily through the introductions to the Subject Indexes. Thousands of copies of these introductions have been provided by CAS. Mastery of the basic rules for systematic nomenclature is a most important step in learning to use the *CA* Subject Indexes. Its teaching is a complex subject in itself which cannot be included here.

So far teaching techniques used for the traditional printed publications have been reviewed. Of more interest at present is the training of chemists in the handling of the new computer-based services which are becoming such an important facet of all modern information processing.

There are two fundamental facts in regard to the transitions through which CAS is going. First, the conversion of the handling, editing, and publishing steps to computer-based operations is being carried out so that accessible search files will be available in addition to the traditional publication services. From these files CAS will be able to select material in any form useful to the subscriber. Second, in this conversion to a computer process, CAS is introducing the concept of a unified data base in which all material derived from the original literature is compiled and from which all CAS publications and services will be withdrawn.

These are important facts which must be heard and understood by every chemist who expects to use the Data Base for his advantage. The teaching problem now continues to include all of the knowhow and horse sense of the intellectual effort mentioned above, plus the requirement of at least a listening knowledge of the computer Data Base and how to address it. CAS has attacked this larger teaching problem on several fronts.

The presentation of technical papers on developments in the growing computer system was enhanced to the point where a few seminars were almost entirely CAS-staffed. Recognized scientific meetings were employed, and continue to be employed, as a media to present the results of CAS continuing research and development in information science. In addition, CAS initiated the custom of an open forum at every national ACS meeting. Here is presented some aspect of the new computer system, and an opportunity is provided for users to ask questions, make suggestions, and criticize. The forum, now in its fifth year, has averaged over 100 attendees and has provided at least a substitute classroom for instruction in the developing computer systems.

With the advent of the computer services and the need to disseminate information about them quickly, CAS staff members prepared articles of general interest for use in the more widely-read trade publications. Papers by Tate[14] on the CAS concept of a unified computer-based information system and by Davenport[15] represent such teaching documents.

The new CAS computer-based services are actually teaching tools in themselves. In 1961, CAS introduced *Chemical Titles* (*CT*), the world's first completely computer-produced publication. In 1965, *Chemical-Biological Activities* (*CBAC*) was first published; in 1967, *Polymer Science and Technology* (*POST*) was initiated, and in 1968, *Basic Journal Abstracts* was developed. These later services are much more sophisticated than *CT* in their use of the computer, and each represents a new development in computer handling of scientific data. These publications and their accompanying magnetic-tape services are pilot plants by which CAS is learning to produce the *CA* of the future. Each employs a different experimental technique which is being investigated before *CA* can be produced entirely through the computer. Each is a teaching tool by which the subscriber can learn in his own hands how to use the future *CA*.

Computer storage and retrieval of scientific data is still very much an art. CAS published its four-millionth abstract on August 5, 1968, and will produce another million abstracts between now and 1971. The techniques for handling this huge mass of data by computer can be learned only by actual experience. The man-machine problems inherent in processing and retrieving more than one quarter of a million abstracts per year can be solved only through experimentation with gradually larger and

larger data bases. Therefore, CAS is offering its subscribers the opportunity to learn by handling small subsets of this large data base. The special publications, which are computer-produced and available on tape, represent from 5 to 10% of the annual total *CA*, and provide an excellent opportunity to learn the techniques of data base manipulations, particularly searching. In this sense, they are most certainly teaching tools.

Because CAS is handling a greater mass of data than any other information processor and because what CAS learns about computer processing will be of assistance to all who follow, no matter what data they may handle, about one half of the CAS research and development effort has been supported by the National Science Foundation. There has also been support from the Department of Defense, the National Library of Medicine, the Food and Drug Administration, the National Cancer Institute, and the U. S. Office of Education. Reports of such contract research are now being made available to the public. They comprise another form of teaching texts valuable to those who wish to develop similar information systems.

A further short, rapid teaching technique is the two-year old CAS traveling seminar. Conducted by three or four senior members of the staff, this team has presented one-day seminars in ten large-population areas of the United States and Canada. Like the open forums, these seminars create an atmosphere in which the interplay of processor and user yields valuable information not only as to how to use the developing computer system, but also suggestions as to how the system may be improved to make it more useful.

As the CAS integrated man-machine system has emerged, it has been accompanied by the realization that the staff alone cannot develop all of the systems necessary both to prepare and to disseminate completely the entire data base. Accordingly, CAS has determined that its staff will concentrate on developing the data base and evaluating general search procedures, and will leave to others the task and the opportunity to develop the specific search systems and new concepts in the distribution of the data. This decision was prompted also by recognition that the data base for all of chemistry will be so large that with present computer hardware, the total store of data cannot be examined by many people at one time. In recent months, these facts have led to the concept of local information centers. Academic or industrial organizations will have subsets of CAS data, plus data from other information processors. Such centers will integrate several data systems and produce summaries in printed or tape formats on subjects of interest to local users.

This concept of local information centers presents CAS with a tremendous potential teaching aid. The most successful of such centers to date is at the University of Nottingham in Great Britain. There the British Chemical Society has established a research unit to learn how to use tape services. The British researchers found almost immediately that the most difficult task in learning to use a computer data base is the construction of proper search profiles. This is the trade name for the questions which a searcher must input in order to retrieve the answers which he seeks. The questions must be framed in language which corresponds to that employed in the data base. It includes a knowledge of computer logic and of the particular data base which is being searched.

The British learned that the best procedure was to train liaison officers at their Nottingham unit and then to send these officers to work with individual laboratory researchers who desired to make searches of the CAS tapes. At present, the liaison officers are working with about 500 British science graduate students, training them in the art of profile preparation and assisting them in running successful searches on the Nottingham data base. The graduate students submit profiles which are run periodically against the data base. Answers are returned to the students who, with the aid of the liaison officers, revise their profiles to obtain more or fewer hits, as their needs indicate. The greatest accomplishment of the Nottingham Unit may well be this training of students in the use of computer services. It certainly provides a new dimension to the CAS program for training chemists in the use of information services.

Information centers similar to that at the University of Nottingham are operating in this country at the University of Georgia, the University of Pittsburgh, and the Illinois Institute of Technology. Others are being considered.

In order to train information-center personnel from overseas in handling the data base, CAS has established an intern program. Qualified interns are accepted as regular CAS employees for a term of not less than one year. During this year, they learn some phase of the CAS operations so that they may return to their center fully prepared to operate this portion of the information system. Under this plan, three interns from Great Britain, one from Sweden, and one from The Netherlands have worked in the CAS offices. Applications of potential interns from other countries are being examined.

Training chemists in the use of the new CAS computer data bases requires individual instruction. Such instructions will be provided through the concept of local information centers to be manned by trained and experienced personnel who have an interest in serving their colleagues in the local area. The philosophy of learning to search the computer store is not a great deal different from learning to search the large printed index. As was emphasized earlier, proficiency in searching comes only through much practice. This is equally true of the more complex computer-stored index and data base.

On the first page of every regular *CA* issue, down in the lower left-hand corner, there appears a statement: "It is the careful endeavor of *Chemical Abstracts* to publish adequate and accurate abstracts of all scientific and technical papers containing new information of chemical interest..." No matter what the service or its form in the future, CAS expects to adhere to this 62-year pledge for a "careful endeavor." With the generous help of many colleagues, CAS attempts to carry out this "careful endeavor" also in the training of chemists.

LITERATURE CITED

(1) Crane, E. J., "The Use of Chemical Abstracts," *Advan. Chem. Ser.* 4, 3 (1951).

(2) Crane, E. J., A. M. Patterson, and E. B. Marr, "A Guide to the Literature of Chemistry," 2nd ed., 397 pp., Wiley, New York, 1957.

(3) Crane, E. J., "Words and Sentences in Science and Industry," *Science* **86**, 549–53 (1937).
(4) Crane, E. J., "Abstracting and Indexing," *J. Chem. Educ.* **20**, No. 11 (1943).
(5) Crane, E. J., "CA Today," *Chem. Eng. News* **26**, 2190 (1948).
(6) Crane, E. J., "Scientists Share and Serve," *Chem. Eng. News* **29**, 4250 (1951).
(7) Crane, E. J., "Chemical Abstracts Yesterday, Today, and Tomorrow," *Proc. Chem. Soc.*, p. 334, Dec. 1957.
(8) Bottle, R. T., "Use of the Chemical Literature," 231 pp., Butterworths, London, 1962.
(9) Burman, C. R., "How to Find Out in Chemistry," 220 pp., Pergamon, Oxford, 1965.
(10) Mellon, M. G., "Searching the Chemical Literature," 40 pp., ACS, Washington, D. C., 1964.
(11) Mellon, M. G., "Chemical Publications," 4th ed., 324 pp., McGraw-Hill, New York, 1965.
(12) Gould, Robert F., Ed., "Searching the Chemical Literature," *Advan. Chem. Ser.* **30**, 326 pp. (1961).
(13) Bottle, R. T., "Use of The Chemical Literature," 231 pp., Butterworths, London, 1962.
(14) Tate, F. A., "Progress Toward A Computer-Based Chemical Information System," *Chem. Eng. News* **45**, 78–90 (1967).
(15) Davenport, W. C., "CAS Computer-Based Information Services," *Datamation*, pp. 33–8, March 1968.

Retrieving Chemical Information with *Index Chemicus**

G. S. REVESZ and ANDREA WARNER
Institute for Scientific Information, 325 Chestnut St., Philadelphia, Pa. 19106

Received October 23, 1968

By communicating in the international language of structural diagrams, *Index Chemicus* provides for rapid scanning and retrieval of information about chemical compounds via subject, author, and molecular formula indexes. In addition, data on instrumental analyses and proposed biological and other uses of compounds are presented. *Index Chemicus* has been extended to provide for computer retrieval and identification of compounds via line notations of structural formulas, and coverage of newly reported reactions of old compounds is now provided.

The basic objective of *Index Chemicus* is to present information about new chemical compounds and reactions in a simple and quickly understandable form, and to make this information easy to retrieve through a variety of indexing methods.

Published weekly, *Index Chemicus* contains an average of 350 abstracts which describe about 3000 new compounds, gathered from the world's chemical literature. Averaging 80 pages per issue, the publication functions as both a secondary journal and a unique registry of new chemical compounds.

Index Chemicus, IC, Index Chemicus Registry System, ASCA, OATS, and *ISI* are registered trade marks of the Institute for Scientific Information. Articles published in *IC* are selected from a basic list of 180 regularly scanned journals and meet one criterion: each contains newly synthesized or elucidated chemical compounds. If an article meets this criterion, it is abstracted, regardless of the country of origin, language, or specific field of chemistry.

The journals regularly processed by *Index Chemicus* are scanned page by page by a team of chemists as soon as they appear in print, some even as they appear in page proofs.[1] These are primary chemical journals, which probably include close to 99% of all published new compounds.[2]

In addition to these regularly-indexed journals, the *Automatic Subject Citation Alert (ASCA)*, another service of the Institute for Scientific Information, provides articles from 2000 additional journals in allied fields. This input yields the remaining few compounds indexed. The articles from a particular journal appear in *Index Chemicus* between 35 and 45 days after the journal is received.

Each abstract of *Index Chemicus* contains the following information:

a. Bibliographic data (author's name and affiliation, mailing address, journal citation, received date, and language)
b. Summary (if supplied by author)
c. Identification of new compounds by structural diagrams and molecular formulas; reaction flows are included when relevant
d. Proposed uses of compounds, if tests and accompanying data are presented
e. Instrumental data and indication of new reaction, when appropriate
f. Author's approval of abstract

In addition, a monthly index is provided, which is cumulated semiannually and annually for retrospective searching by:

g. Author
h. Subject
i. Molecular formula
j. Journal
k. Rotaform[3]

Optional services available to the reader include

l. Magnetic tape search
m. Tearsheet or reprint of original journal article
n. Magnetic tapes with structural diagrams in line notation form

*Presented before the Division of Chemical Literature, 156th Meeting, ACS, Atlantic City, N. J., September 13, 1968.