DOCUMENT RESUME

ED 040 731 LI 002 051

AUTHOR Holm, Bart E.
TITLE Modern Techniques for Searching the Chemical Literature.
PUB DATE Apr 70
NOTE 7p.; Presented at Conference on Use of Modern Methods of Handling Chemical Information in The Education of Chemists, April 2,3, 1970, Washington, D.C.

EDRS PRICE EDRS Price MF-$0.25 HC-$0.45
DESCRIPTORS *Chemistry Instruction, Chemistry Teachers, *Information Dissemination, *Information Processing, *Information Services, *Information Utilization

ABSTRACT

The chemists' information needs are for current awareness, selective dissemination, and retrospective search services, of research, development, engineering, production, and marketing information located internally or externally, and contained in journals, patents, theses, reports, data files, information services, and from people. This paper is an overview of approaches to the processing of chemical information including new techniques for handling structures, concepts and data. These methods are available, many are inexpensive, and they are widely used in industry and the government. They can also be helpful to chemistry teachers and students. (NH)

Conference on Use of Modern Methods of Handling Chemical Information
in The Education of Chemists

April 2, 3, 1970
Washington, D.C.

MODERN TECHNIQUES FOR SEARCHING THE CHEMICAL LITERATURE

Bart E. Holm
Development Department
E. I. du Pont de Nemours & Co., Inc.

INTRODUCTION

Beilstein once said, "I read everything. I place it where it belongs." You can't do this any longer except in narrow fields. You are blessed or plagued with an abundance of published and unpublished information (1*). This is an historical continuation of a growing wealth of knowledge waiting to be put to use (2). Chemists are fortunate, for of all the disciplines chemistry is unique with a tradition of information excellence and a tradition information organization. Although you have growing information resources, you do have new and improving mechanisms for managing these resources. I will state the types of services a chemist requires and introduce the modern techniques and formats used in these services. The speakers and the demonstrations to follow will show, in detail, how these methods are being widely used to help the chemist. Many of these you can apply when you get back to the campus, many are inexpensive, all are useful. We hope you will see ways to advance the educational process through these techniques and services.

CATEGORIES OF NEEDS AND INFORMATION

The information needs of chemists can be stated as the need for current awareness, selective dissemination, and retrospective search services, of research, development, engineering, production, and marketing information located internally or externally, and contained in journals, patents, theses, reports, data files, information services, and from people (3). Current awareness is continuous professional education through reading journals, attending meetings, and informal communications with colleagues. Selective dissemination is receiving information related to your interests from the remaining published literature beyond that scanned. Retrospective search is looking for information to answer a specific need. Chemists have traditional and new services in each of these categories.

Chemical information can be categorized as dealing with structures, data, and concepts.

***Figures referred to in this text have been omitted, due to marginal legibility of the illustrations. (ERIC/CLIS note)**

Structures

Most of the effort in the information processing field has been in chemistry with much of that effort on improved ways of manipulating chemical structures. Chemists use chemical structures to communicate information about a chemical compound. Chemists need access to structural information, and the structure is a discreet entity amenable to organization.

Chemists commonly represent chemical compounds with pictures, an excellent method for one chemist to communicate with another using a blackboard. Each understands the meaning of the message. Unfortunately, the diagrams are not pronounceable, the diagrams cannot be ordered, and the multidimensional representations are difficult for printers (4).

A number of nomenclature systems are in existence, each assigns one correct name for each compound. Nomenclature serves well for oral communication and for printing but nomenclature is not very effective for class searching because the main structural part carries many different names, depending upon the structure of the rest of the compound.

Beilstein Classification is a common way of identifying chemicals but it is difficult to search for subcode information unless you know the major classification.

Pictures, nomenclature, and classification have a place, but each has drawbacks for organizing large collections for indexing and searching. The major emphasis in recent years has been on fragmentation, notation, and topological coding (5).

In fragmentation a compound is represented as a composite of its major structural features. Paul Craig will describe fragmentation coding. With notation the structure is represented by a single line of symbols. Al Smith will describe chemical notation systems. In topological coding the structure is represented by a unique array of symbols in a compact format for storage and search. Fred Tate will describe the Chemical Abstracts Registry System which includes topological coding. Some of the demonstrations will show you how these methods have been applied.

These three approaches are major advances enabling chemists to manipulate large collections of chemical structures to locate specific structures and to bring together groups of structures. The Committee on Chemical Information in 1969 published "Chemical Structure Information Handling - A Review of the Literature 1962-1968," Publication 1733, National Academy of Sciences, Washington, 1969.

Data

Data are numeric or quantitative notations helping to describe a subject more precisely. Chemists are interested in melting points, boiling points, molecular weights, and specific gravity. These identifiers are easy to list and manipulate manually or by machine. One area, analytical data often includes instrumented techniques yielding indefinite curves or figures for compounds and mixtures. For identification these must be compared with the curves or figures of pure standards. Searching data by computer will be discussed by Carlos Bowman tomorrow followed by a demonstration of infrared spectra searching by Duncan Erly.

Concepts

Concepts cover qualitative information in chemistry as contrasted with data and structures. Concepts interact and overlap and are not mutually exclusive (6). Ablation, polymerization, extrusion, and spinning are all concepts and part of the chemical literature. Concepts have their own forms of presentation and manipulation. Let me review some of the new approaches.

ACCESS TO QUALITATIVE INFORMATION

Current Awareness

Chemists read journals to learn the details of new developments and to keep current in their field. Most professionals do not try to scan all the journals that might have articles of interest. Those that try to scan many journals usually end up with tables in their office, den, or bedroom stacked with unread magazines. Subscribe to and regularly read a few core journals. Use a current awareness service to alert you to other material. Of the new services available two are representative. One called Current Contents, Chemical Sciences, gives title information by reproducing the tables of contents of hundreds of foreign and domestic research journals. The weekly publication is extremely fast. The usefulness of the idea has caused many organizations to create their own table of contents publication to cover their journals in their area of interest.

The second service, Chemical Titles, listing titles of articles, is issued biweekly and covers 700 journals. Titles are rotated or permuted to list the significant words in alphabetical order (7). Since the work can be done by computer it offers efficient, inexpensive presentation. The title appears as many times as there are significant words in the title. A list of common words called a stop list is used to prevent alphabetizing on articles and other common words. These permuted listings are called KWIC (Keyword-in-Context). Some KWIC indexes drop the title information that cannot be included within a set number of characters for the line. An advance was made by taking the keyword out of the title and listing it on the left side of the page followed by the full title (8). These systems are called KWOC (Keyword-out-of-Context).

The next level of completeness for current awareness beyond scanning titles would be to receive abstracts. The total abstract publications, Chemical Abstracts, Biological Abstracts, and Physics Abstracts are seldom used for current awareness today because of size. The approach by these organizations, is to produce subsets of the total in narrow fields. Chemical Abstracts is published in sections. CAS also offers Basic Journal Abstracts, Chemical-Biological Activities, and Polymer Science and Technology. Index Chemicus, published weekly by the Institute for Scientific Information emphasizes synthesis, isolation, identification and/or biological activity of new chemical compounds. Other abstract services are cover patents. These are all examples of published current awareness services. You will see examples of some of these in the demonstrations.

Selective Dissemination of Information

Keeping up with the literature by current awareness techniques is often too big a task for an individual. Selective dissemination of information, usually called SDI, is the process of sending specific information to a person. Good librarians have done this for years. As they review the material coming into their collection, they send items to those known to be interested. People fortunate enough to have executive staff assistants have the same job performed for them. Now some services are available to help you and me.

Each person participating in an SDI service prepares a list of terms, called a profile, describing his area. As material comes in, the profiles are matched against the indexing terms, the titles, the authors, the references, or the abstracts to look for matching items. The person receives notification of newly received literature that matches his requirements.

In the most common form the notification consists of two punched cards (9). The first card contains a title and sometimes an abstract. The second card is a form to request the full article, and also gives a way to comment on how close the item matches the request. Systems adjust the profiles based on this feedback.

Some publishers offer SDI services to individuals or will provide magnetic tapes for processing by your organization for local SDI services. CAS offers magnetic tapes for Chemical Titles, Chemical-Biological Activities, CA-Condensates, and Polymer Science and Technology. Engineering Index offers a similar data base prepared from 3500 engineering publications. Excerpta Medica Foundation provides magnetic tape files covering 3000 biomedical journals. ASCA, Automatic Subject Citation Alert, from the Institute for Scientific Information processes profiles against the input to the Science Citation Index. The user supplies a list of authors or cited references and receives as

a product all newly published work citing his stated references (10). In essence ASCA, or the Science Citation Index brings the researcher forward in time by listing those recently published documents which cite to one of interest to him. In addition to authors and references, profiles can consist of combinations of words, word stems, word phrases, organizations, and journals.

Current awareness and SDI services offer practical, economical, methods to assure that you are alerted to newly published work from all over the world. Read just a few core journals. Use current awareness and SDI for the rest.

Retrospective Search

You have heard how to keep current with the published literature. Some items you will read and discard; some you will keep. As soon as anyone accumulates a few thousand of anything, except money, he may have a problem finding what he needs when it is required. Here again, there are simple ways to organize your personal collections to find any item easily (11).

First consider how you will search your files. Decide whether you will use single entry or multiple entry files. The admissions office keeps its records by the names of students. A similar system can be used in chemistry for a file on physical properties of chemical compounds if the only questions asked are to supply property values of specific compounds. In these cases pick your file subjects and set up the familiar classification system.

In science, frequently this does not occur. The need might be to locate all chemical compounds with a particular property value. A multiple entry system can handle multiple subject searches.

Traditional systems describe an item by a term or two. Index more deeply. Describe the articles in your collection by five to ten subject terms (12). Fully describe each item in order to provide flexibility in finding what you want quickly and economically.

The terms you select must be organized for consistent input and search. A modest amount of vocabulary control can assure that information is not lost because you indexed it one way and searched for it in another. Be consistent. Use CA nomenclature. The Condensed Chemical Dictionary is another good reference. If your subject area is extensive and complex consider formalizing your vocabulary by recording synonyms, generic levels, and related terms (13).

A variety of storage forms are available for the index. First, give each article or item in your collection a sequential file number and store the items in order. Then set up a record for your index. Here is one form (14). A card is used for each term. Record the item number on the term card that applies. In searching select the terms of interest and look for matching item numbers. In the example report 100 discusses "ablation of plastics in heat shields of space vehicles".

The Committee on Chemical Information has a file of over 300 articles on recent significant developments in chemical information processing. Abstracts for each are kept in serial order. The storage form for our index is an optical coincidence deck (15). Each card is a term. Item numbers are recorded as holes in the card and answers to questions are found by overlaying term cards. Samples of these two storage forms and one other will be shown to you tomorrow in a demonstration. Any of these methods can assure quick, complete retrieval from your files.

Microforms

Files can become voluminous. Collections of journals, abstracts publications, and patents can grow larger than the space available. Microforms have come into active use to reduce storage requirements, provide rapid access, and are an inexpensive way to provide multiple copies of whole documents. The common microforms are microfiche, aperture cards, and roll film in cartridges (16).

A microfiche is sheet microfilm about 4 x 6 inches commonly containing 5 rows of 12 images and a header strip for document identification. Government reports are available as microfiche for $0.65 versus $3.00 for full size hardcopy. The Committee on Chemical Information has its document collection on microfiche as you will see tomorrow (17).

Aperture cards are tabulating cards with a hole in which the film is placed. Normally one frame is included in the card although some systems hold up to eight frames. The Department of Defense stimulated the use with the ruling that all engineering drawings submitted for their contract work must be on aperture cards meeting their standards. U.S. patents are now available on aperture cards (18).

The most familiar and economical microform is 16mm roll film. For high usage the film is stored in cartridges which simplifies threading in reading equipment. A 100 foot reel of 16mm film usually contains 2500-3000 8-1/2 x 11 inch pages of material. The American Chemical Society offers many of their publications in this form including ACS journals and Chemical Abstracts. Some of these products will be shown to you this afternoon (19).

Chemical Information Resources

Chemical information is widely available in the United States. The American Chemical Society is active in providing primary and secondary services both as a wholesaler and a retailer. Of the commercial services in chemistry, the Institute for Scientific Information is the most extensive. A number of their products have been described and you will see others during a demonstration tomorrow.

The BioSciences Information Service publishes a variety of products in their field including Biological Abstracts, and a KWIC index called Biological Abstracts Subjects in Context.

The National Library of Medicine here in Washington is a reservoir of published literature in the medical sciences. Each month an abstract journal, Index Medicus, is published. All incoming journals are indexed for storage and retrieval for their computer system called MEDLARS.

The United States Government has many other information services of value to chemists. A notable resource is U.S. patents. The U.S. patent office has issued over 3.5 million patents, a rich source for chemists. The patent office has concentrated in recent years in improving their method for supplying copies of patents both as hardcopy and in microform. They have done little to improve the intellectual access to the content of patents. Commercial services fill this gap by offering abstracts and indexes to U.S. and foreign patents. In the United States, Information for Industry publishes the Uniterm Index to Chemical Patents and has magnetic tape services. In London, Derwent Publications inaugurated a Chemical Patents Index in January, 1970 providing abstracts, microfilm, and eventually indexes to worldwide chemical patents.

Summary

This has been an overview of new approaches to the processing of chemical information. Remember that a chemist needs a balance of current awareness, selective dissemination, and retrospective services. New techniques for handling structures, concepts, and data are available to help him in his work. New publications, formats, and services permit him to choose the inputs to his problem solving procedures. These are all tools to help the chemist do his work more effectively (20). The methods are available, many are inexpensive. They are widely used in industry and in the government. Can these methods help your faculty and your students, now and in the future? We are convinced that they can.