Indexing and Abstracting Chemical Information: The View of Two Industrial Chemists

PAULA B. MOSES* and LEE E. NELSON

Technical Information Services, Dow Chemical Company, Midland, Michigan 48640

Received December 5, 1983

The proliferation of chemical literature and widespread availability of on-line bibliographic databases have significantly affected traditional indexing of scientific literature. The strengths and weaknesses of current chemical information indexing are explored. Specifics from the published literature and a proprietary database are provided. The developing trends of end-user searching, full-text databases, and graphics technology are explored in relation to types of information and data that scientists expect to access rapidly and efficiently. These trends are brought into perspective by the "megatrends" of an information society that demands the duality of "high tech" and "high touch". The future of indexing and abstracting chemical information is discussed. Concomitantly, the developing trends in information storage and retrieval are brought into focus.

In order to appreciate the role of chemical information storage and manipulation, we must first explore traditional indexing. What is it and why do we need it? This must be placed in perspective by the changing needs and direction of information retrieval.

Indexing of chemical information began more than 75 years ago when pioneers in this field recognized the growing need to simplify access to the world of chemical literature. The philosophy of indexing and sophistication of indexing tools developed dramatically during the next 50 years. As Sputnik rocketed into orbit in 1957, the "Information Age" began. Mainframe computers emerged, and the volume of information being generated escalated. During this era, indexing and abstracting thrived because they are an essential bridge between information generation and information utilization. In another perspective, indexing and abstracting are critical ingredients in efficient utilitization of published information.

Indexing involves the intellectual analysis of a document: picking it to pieces to discover its key ingredients. Equally as important in indexing is putting the pieces back together in the form of an abstract. Stating the essence of the work is crucial. Capsulization is an integral part of this process.

Because of the basic ambiguity of language and the desire for precision in information retrieval, standard indexing vocabularies evolved. Many papers and presentations have been prepared on the advantages and pitfalls of indexing. From a chemist's viewpoint, several strengths of indexing are important to note. Indexing, with its standardized terminology, provides access to chemical structures and technology. It allows retrieval of numerous documents on the same topic when authors have used different terminology to describe the same technology. The same applies to retrieval of information about chemical substances. Indexing, when properly standarized and cross-referenced, allows us to locate documents on the same substance or ring backbone whether authors use Chemical Abstracts Service nomenclature, tradenames, or trivial names. These retrieval tasks would be much more difficult and time consuming without standardization and cross-referencing in indexing.

The quantity of published chemical literature is growing each year. Chemical Abstracts Service alone cited more than 500 000 documents in 1982. Indexing and abstracting this wealth of knowledge is big business. The information suppliers deserve a round of applause for making these valuable resources retrievable by the scientific community.

Indexing and abstracting policies are constantly being evaluated and improved. Three specific areas that need to be addressed in this light are depth of subject modifications, experimental data, and material and physical properties. Indexing cannot be so exhaustive that every topic discussed in a document is reflected in the index, yet detailed indexing enhances retrieval of pertinent documents. The volume of chemical literature is overwhelming. Scientists need to be able to specifically pinpoint pertinent documents without wading through those of peripheral interest. More thorough and precise indexing would help achieve this need.

The Dow Corning Corp. has recognized this need and generates supplemental indexing of published literature pertinent to their business. Some categories covered in supplemental indexing include applications, reaction conditions, material properties, toxicological data, and polymeric substructures. Applications include product use or market area. Additional indexing emphasis is placed on reaction conditions and material properties. Specific toxicological data are indexed. Polymers are encoded for structural searching.

The Dow Chemical Co. does not augment indexing of published information. However, proprietary technical reports are indexed extensively. Bibliographic information and indexing terms for more than 200 000 Dow technical reports are searchable in an on-line database. Indexing philosophy is based on Chemical Abstracts Service policies, but it is customized for company needs. The indexing is more thorough and precise, with greater emphasis placed on polymer characteristics.

Trends in both the Dow Corning Corp. and the Dow Chemical Co. automated information systems include use of text-processing software and machine capturing of data at the source. Computer-security packages are used to reduce the chance of confidential information falling into the wrong hands.

Before addressing the changing needs of information retrieval, we must define scientists' perception of their information needs. In a nutshell, they want, first, an efficient mechanism for surveying the world of chemical literature. Scientists want an easy but precise way to search the literature. A minimal amount of training is a critical aspect. If it takes 1 week to learn to use the system and several hours a month to stay abreast of changes, they will not use the system. Second, they desire rapid access to specific data of interest, not references to articles and, often, not even full-text articles, but actual data. They do not want to be hampered with dwindling library subscriptions to journals or books, nor do they want the delay of obtaining interlibrary loans of articles. In short, they want what they want when they want it. Last, but most important, they want confidence in the completeness and accuracy of the search. Halfway through a costly research project, the scientist does not want to discover that his idea was explored and patented 2 years earlier by someone else.

John Naisbitt, a social forecaster, recently gained considerable acclaim for his best selling book *Megatrends*. This book analyzes the major societal, political, and economic currents shaping our future. Clearly, some of these "megatrends" bear

0095-2338/84/1624-0189$01.50/0

directly on the information services industry. The first megatrend affecting this topic is that Americans continue to believe they live in an industrial society, despite the obvious fact that the economy has changed and they now live in one based on creating and distributing information. Many are employed as professionals. Professional workers are almost always information workers—chemists, lawyers, teachers, computer programmers, and others. In an information society like ours, capital is a strategic resource, but information is the key. Access to this information is crucial.

The second megatrend is the duality of a "high-tech" and "high-touch" society. Naisbitt means that each technological application brings with it a need for a compensatory human response. There are two key points: we live in an information-based economy with rapidly escalating technology, yet society demands that technology be balanced with human responses. This represents high tech and high touch.

These megatrends already affect the direction of infromation creation, processing, and distribution. We see it in development of end-user searching, full-text searching, and graphics technology.

Why does the user want to search the literature himself? We believe it is not because of a lack of trust in the information intermediary. Rather, the fancy computer databases represent high tech and the user wants personal interaction with them or high touch. He feels that this personal interaction or manipulation of data will enhance creativity and innovation. The same factors come into play with full-text searching. Additionally, full-text searching expands access to nonindexed information. It speeds access to information by eliminating the need to first locate an abstract and then find the document of potential interest. Full-text publishing additionally can provide more currency to information dissemination since the material does not need to be indexed or abstracted.

Graphics technology provides substructure-search capabilities. It allows data manipulation, such as modeling simulation, and statistical and pictorial representation of data. It is high tech. When a scientist sits at a graphics terminal and uses a light pen or tablet to build a model of interest, it represents high touch. Graphics also enhances electronic publishing and is a crucial link in acceptance of full-text publishing.

In this information-based society with an emphasis on high tech and high touch, will traditional indexing and abstracting survive? Yes—at least for awhile; indexes and abstracts are an important means of communication. They focus on the essence of the work reported in an article. Many scientists are not interested in reading all articles in their field of expertise but choose to stay abreast of technology advancements by reading the author's abstract or one prepared by an information service.

Anyone who has had the opportunity to search a full-text database appreciates the efforts that go into and the importance of indexing. Unquestionably, there is a time and a place for full-text search capabilities, but the need for information manipulation, such as indexing and abstracting, is more important as the volume of published information increases.

Traditional indexing and abstracting will change. The business of information storage and retrieval is growing as rapidly as a snowball rolling down a mountainside. This snowball is increasing dramatically in size and is gaining momentum as it gathers snow and falls farther down the mountain. This momentum brings with it more and more information users. These users, whether they are scientists, government regulatory personnel, or information specialists, are becoming increasingly sophisticated. Their information needs are multidisciplinary. Their time frame for needing answers is short. They want access to more specific information and to data. Some examples include toxicological data, precise testing methods and procedures, reactivity of specific atoms, and the ability to search polymers substructurally. This information needs to be rapidly retrievable. Indexing needs to be expanded and more precise. Technological advancements will help index and abstract more efficiently and effectively.

Types of information requests will continue to vary from bibliographic records containing index terms, to full text, to actual data. Full-text publishing, complete with search capabilities, will not, within the next few years, preclude the need for indexing and abstracting. Searchable databanks with creditable, verified data such as an LD_{50} or specific physical property data will not preclude the need for indexing and abstracting. The ability to graphically search chemical structures is a grand technological advancement but is no substitute for searching index files. All of these retrieval avenues are important. They complement indexing and abstracting. But indexing is the necessary bridge between information generation and effective information utilization.

A Practical Approach to the Use of Literature Early in a College Career†

EUGENE A. KLINE

Department of Chemistry, Tennessee Technological University, Cookeville, Tennessee 38505

Received November 21, 1983

The use of literature in freshman and sophomore chemistry courses can be implemented in the laboratory very easily, especially in Organic Chemistry. Slight alterations of the prescribed synthetic procedures will encourage the student into literature to seek all kinds of necessary information and will play an important part in teaching the need for literature searching.

The importance of professional literature for chemists makes it most worthwhile to examine various ways of stimulating its use by undergraduate chemistry majors in the first 2 years. However, it has to be noted that differences in emphases among departments, abilities of students, and availability of literature demand that several options be considered.

Ideally, the use of chemical literature by undergraduates should commence as soon as possible. There are even Freshmen who can and should be trained to do serious literature searches, and ways should be developed for those freshmen who need special elementary assistance to acquire the knack of such searches. For example, handbooks and

† This paper was presented at the Symposium on Chemical Literature and Information Retrieval in the Chemistry Curriculum, 184th National Meeting of the American Chemical Society, Kansas City, MO, Sept 15, 1982.

0095-2338/84/1624-0190$01.50/0