Award Symposium on "Contributions of the Division of Chemical Literature (Information) to the Chemical Society"*

Milestones in Chemical Information Science†

HERMAN SKOLNIK

Hercules Incorporated, Research Center, Wilmington, Delaware 19899

Received April 6, 1976

Chemical information science is examined retrospectively, introspectively, and prospectively as a discipline of chemistry. Early developments briefly discussed are the library, classification concepts, scientific societies, scientific journals, chemical nomenclature, abstract journals, and indexing systems. These early developments constituted the heritage on which major contributions have been based since the 1940's, and which are explored, such as chemical information science as a career, edge-notched punched and optical coincidence cards, notation systems, nomenclature, indexing systems, computers, and the information industry.

Our commemoration and celebration this week of the centenary of the American Chemical Society, which was founded on April 6, 1876, provide a most appropriate opportunity to examine our discipline of chemistry, chemical information science, retrospectively, introspectively, and prospectively.

To begin with a prediction, I think the future will look back on the period 1943 to the present as the Golden Age of chemical information science. We think of a Golden Age, such as the Greek civilization of the time of Aristotle, Plato, and others, not as a birth but as a blossoming, as a period of high and meaningful activity, as a period that shapes the future. The Golden Age of chemistry occurred in the 19th century, when chemistry became a science by the construction of a language, a system of symbols and nomenclature, on which a cumulative literature could be based. Every Golden Age I can think of, and there are many, has been marked by the formation of a substantial information and knowledge base. Yet information science was not an object of interest in the historical evolution of civilizations. This is not to say that the accumulating information and knowledge bases were without tools and systems for collating, interpreting, relating, and even retrieving the results of scholarship. These activities were a part of scholarship, something that everyone did rather than specialize in. This was markedly true during the 19th century in the historical evolution of chemistry into a science, an educational curriculum, and an industry.

Information science and technology is as old as the written word. It is associated closely with the evolution of human communication and every aspect of education and training. Some of the early milestones in information science and technology were:

1. **Writing**, the outstanding invention of our prehistorical ancestors, is estimated to have occurred around 3500 B.C. Writing was an essential factor in the continuing evolution of civilizations.

2. **Written records** on paper for the preservation of information and knowledge came into existence probably around 600 B.C. in Greece and somewhat earlier in India and China.

3. The **Library**, for the remembrance of all things past, dates from about 2000 B.C. for the collection of clay tablets, the "books" of the Babylonian civilization. The earliest large collection was that of Alexandria, started by Ptolemy I, the king of Egypt, about 300 B.C., which at its height contained about 600 000 parchment scrolls (equivalent to about 100 000 books of today). The office of librarian was one of the highest that the king conferred.

4. Concepts of **definition** and **classification** as developed by Aristotle (384–322 B.C.).

5. Invention of **printing** about 1450 marked the democratization of communication by the printed word, with a flood of pamphlets and books becoming readily available to the general public. This invention also marked the beginning of personal and public libraries throughout Europe.

6. **Encyclopedias**—one of the first attempts to gather all knowledge into one work was that of Aristotle. The first attempt to arrange knowledge in a given order was made by Vincent of Beauvois in 1260. Francis Bacon's "Novum Organum", published in 1620, was the first to classify information. The first significant encyclopedia in chemistry was Mocquer's four-volume "Dictionnaire de la Chymie" (1778).

7. **Scientific societies.** Although the "Accademia del Cimento", founded in 1657 in Florence, Italy, is credited with being the first, the Royal Society, founded in 1662 in London, England, is the oldest that is still in existence and the first to have a publication, the *Philosophical Transactions*, beginning in 1665.

8. **Scientific journals** slowly evolved from newspapers and correspondence between scientists. The *Chemisches Journal* (1778–1784) is thought to be the first chemical journal; it was subsequently named Crell's *Chemische Annalen* (1784–1803). *Annales de chimie* (France), first issued in 1789, is the oldest chemical journal still being published. The most important journal during the 19th century was *Justus Liebig*'s *Annalen der Chemie*, first issued in 1832.

9. **Treatises.** Gmelin's "Handbuch der anorganischen Chemie", first issued in 1817, is the oldest and still one of the most important reference works for inorganic chemistry. Its counterpart in organic chemistry is Beilstein's "Handbuch der organischen Chemie", which was first issued over the period between 1881 and 1883.

10. **Chemical nomenclature.** The first important landmark in chemical nomenclature was the 1789 report, "Methode d'une Nomenclature Chimique", written under the combined

* The paper by K. Loening, "Status of Chemical Nomenclature and Potential Future Trends", is not included here. It will appear in the February 1977 issue of this journal.

† Division of Chemical Information Award Address presented in the Symposium on "Contributions of the Division of Chemical Literature (Information) to the Chemical Society" at the ACS Centennial Meeting, New York, N.Y., April 6, 1976.

effort of Morveau, Lavoisier, Fourcroy, and Berthollet. Chemical nomenclature has been evolving ever since.

11. **Abstract journals.** These were initiated by chemical societies. The French Chemical Society introduced its *Bulletin de la societe chimique de France* in 1858; the German Chemical Society published abstracts in *Berichte der deutschen chemischen Gesellschaft* from 1868 to 1896, then took over *Chemisches Zentralblatt* in 1897; the American Chemical Society published abstracts in its *Journal of the American Chemical Society* from 1897 to 1906 and then established *Chemical Abstracts* in 1907.

12. **Special libraries and librarians.** From the time of the great Alexandrian Library until relatively recently, librarians were scholars. The change from scholarship to service occurred in 1876 with the formation of the American Library Association (ALA). Orientation of librarians to subject areas, such as chemistry, in the service of business or industrial organizations took form with the founding of the Special Libraries Association in 1909. Special librarians in the field of chemistry were the first professionals to specialize in chemical information science.

13. **Book classification systems.** The first widely used method of classifying books in a library was introduced by Melvil Dewey in 1876.

14. **Indexing systems.** The alphabetical subject and author indexes are about as old as the book. They became an art and to some degree a science with the advent of librarianship as an educational curriculum. Other types of indexes arose from the nature of a subject area, such as chemistry. Thus Richter introduced the empirical formula index in 1884 for his tables of compounds. A formula index designed by Jacobson and Stelzner was used in 1898 by *Berichte*, and a formula index designed by Hill for the U.S. Patent Office in 1900 which cited C first, H second, followed by other elements in alphabetical order by their symbols was used by *Chemical Abstracts*. Patterson's ring index, based on the number of rings in an organic compound and kind of atoms in each ring, has been used by *Chemical Abstracts* since 1916.

15. **Correlation systems.** Among the sciences, chemistry has been particularly unique in the number of systems introduced for correlating information. One of the more fruitful early ones was the concept of homology advanced by Dumas in 1851. An outstanding one was the Mendeleev periodic table introduced in 1869.

This brief summary of the milestones in chemical information science and technology from early history until relatively recent times constitutes the basic heritage on which our contributions have been based. In an important sense, we should add a 16th milestone before we consider what we have accomplished in our own times—technical meetings which had programs for the presentation of papers in various areas of information science and technology. The Special Libraries Association (SLA), since its founding in 1909, has had many papers in its annual meetings which have been pertinent to information science and technology and which have been published in its journal, *Special Libraries*. In the 1930's, the American Documentation Institute (ADI) was formed by people in the Washington, D.C., area who were interested in the development and promotion of scientific aids to information processing, such as microforms; this organization, now called the American Society of Information Scientists (ASIS), is the largest in the United States dedicated to this discipline.

Almost from the very beginning of ACS national meetings, papers were presented on various aspects of the chemical literature. The first symposium in an area of information science presented at an ACS national meeting was on "Chemical Libraries and Their Problems" in April, 1919, at Buffalo, N.Y.

By the 1940's, many chemists with unique skills and talents found employment opportunities in the chemical industry as librarian, editor, writer, translator, indexer, abstractor, literature searcher, and others which became subtended under the concept of literature chemist. Many of these literature chemists, trained and educated in the rigors of chemistry, thought of themselves as chemists and as working as chemists, and preferred to be oriented to the American Chemical Society even though they were members of SLA, ADI, and similar organizations. Consequently, these chemists formed in 1943 the Chemical Literature Group within the ACS Division of Chemical Education to plan and organize a variety of sessions for ACS national meetings.

Between 1943 and 1949, the Chemical Literature Group, in addition to general sessions, put on the following symposia:

1. "Technical Library Techniques"—fall, 1943; spring, 1944; and spring and fall, 1946.
2. "Demands for Nonlaboratory Chemists"—spring, 1944.
3. "Chemical Patents"—fall, 1946; spring, 1947.
4. "Petroleum Information Services"—spring, 1947.
5. "Chemical Literature"—spring, 1947.
6. "Research Records"—fall, 1947.
7. "Punched Cards"—fall, 1947.
8. "Nomenclature Problems"—fall, 1947.
9. "Trends in Indexing, Classifying, and Coding Chemical Information"—spring, 1948.
10. "Technical Journalism"—spring, 1948.

By the activities of its members and the programs it scheduled at ACS national meetings, the Chemical Literature Group established literature chemists as bona fide members of the fraternity of chemists in industry and government, and somewhat later in academe. Until then, and even during most of the 1940's, literature chemists were thought of and often classified as librarians. This tendency is still somewhat prevalent because literature chemists do much of their work in libraries and in close collaboration with librarians, and some are indeed chemical librarians.

A large proportion of the members of the Chemical Literature Group during its formative years was oriented strongly to the library science side. Many of these people had received a B.A. or B.S. degree in chemistry or in an allied discipline, such as physics, biology, and home economics, and an M.S. in library science. In library science, the emphasis is on obtaining and storing documents in the form in which they are produced, and good librarians are highly knowledgeable in sources of and the housing of literature. Except for some indexing and adherence to one of three book classification systems, little effort is directed in the library science curriculum to the process of eliciting information from documents so that users can know what information belongs together or can differentiate information generically and specifically.

Another group of members of the Chemical Literature Group during this period more or less drifted into literature or library work without benefit of any scientific or library science background. Because they could not be members of the American Chemical Society, they were called affiliates or associates. Many were highly skilled and talented in areas of importance to chemists, such as technical writers and editors, translators, publishers, and patent attorneys. Many members were trained and educated in chemistry, a quantitative science, but their careers were primarily of a qualitative nature, such as indexer and abstractor, within an established system. Some members trained and educated in chemistry, many through the Ph.D. and with experience in the solution of chemical problems by scientific methods, devoted their energies and talents to finding solutions to the problems associated with the chemical literature. They were concerned with devising new indexing, classification, and communication systems, new tools, and new concepts. These members have been the major contributors to milestones in chemical information science and technology over the past 30 or so years, but not the only contributors, for the other groups of members also contributed milestones since 1943 when the Chemical Literature Group was formed and since 1949 when the Group became the ACS Division of Chemical Literature.

MILESTONES SINCE 1943

Chemical Information Science as a Career. One of the initial and continuing impacts of the Division of Chemical Information on the American Chemical Society and its members was the establishment of chemical information science as a discipline of chemistry with a variety of opportunities for careers in the chemical industry, governmental services, and academe. It provided a forum for the presentation of papers that advanced the art and science practiced by its members, to the point where the American Chemical Society recognized the Division's need for a journal, and introduced the *Journal of Chemical Documentation* in 1961.

Edge-Notched Punched Cards. Edge-notched or hand-sorted punched cards were buried in a series of U.S. patents which issued in 1904, 1907, 1925, 1929, and 1940. They were commercially available in the late 1930's and early 1940's from McBee Co. (Keysort) and Charles R. Hadley Co. (Pathfinder). Publications on their applications appeared in 1939 and 1941, but in journals such as *Journal of Dental Research*, *Metal Industry* (*London*), and *Transactions of the Institute of Welding* (*London*). Awareness of this new tool among chemists, and specifically literature chemists, did not occur until 1945 with the publication of the article, "Punched Cards for a Chemical Bibliography", by Cox, Bailey, and Casey in *Chemical and Engineering News* and with the presentation by Casey and Bailey of the paper, "Punched Card Techniques and Applications", before the Chemical Literature Group at the spring, 1946, ACS national meeting.

At the 1946 ACS fall national meeting six papers described punched card techniques, and at practically every meeting through 1956, papers were presented on various aspects of punched cards.

Punched card techniques were sufficiently novel for the ACS Board of Directors to set up a Committee on Punched Cards in 1946 with financial support into the 1950's. The work of this committee and the papers presented before the Chemical Literature Group and Division resulted in the publication of the book, "Punched Cards", by J. W. Perry and R. S. Casey.

Until the advent of the edge-notched punched card, the 5 × 3 library card was the traditional mode for alphabetical and numerical index files. The traditional 5 × 3 card file required strict arrangement by a systematic order and having in the file as many cards per reference as the parameters of the information system dictated. The primary advantage of punched cards was that one card per reference properly coded could be retrieved by as many ways as there were types of codes, such as author, subject, and reference. The punched card file could be arranged in any order, including random.

In applying punched cards to information systems, various coding concepts were developed, such as the direct, indirect, combination, and superimposed, and the derived mathematics of coding concepts enabled the designer to understand the advantages and disadvantages of the various coding schemes. These two factors proved to be intellectual stepping stones to the use of computers.

Notation Systems. One of the more important milestones in chemical information science over the past 30 years has been the introduction of a variety of notation systems that uniquely and unambiguously represent chemical structures. The first notation system of note was that of G. Malcolm Dyson, which was first presented in 1946 before the Royal Institute of Chemistry in London. Dyson presented a paper on his notation system before the Division of Chemical Literature at the 1947 ACS Spring national meeting. Shortly thereafter W. Gruber, J. A. Silk, and W. J. Wiswesser announced their notation systems.

Interest in notation systems ran high and prompted the National Research Council (NRC) to initiate a program for teams of chemists throughout the United States to evaluate the four notation systems. Many members of the Division of Chemical Literature were members and captains of these teams. The outcome of this program was anything but decisive. Although the Dyson system was favored over the other three, mostly because his manual was the best written, the coding and decoding errors for all four were relatively high. Most importantly, however, the program engendered a widespread knowledge of these notation systems and motivated many chemists to seek new solutions.

In 1949, the IUPAC Commission on Codification, Ciphering and Punched Card Techniques, which prompted the NRC study program, listed a set of desiderata that a new notation system should satisfy. In 1951, the IUPAC Commission adopted the Dyson notation system on a provisional basis, and in 1961 published the definitive "Rules for IUPAC Notation for Organic Compounds", a modification of the original Dyson system.

In the NRC study, some members of the teams discerned merit in the Wiswesser system but complained about the poorly written manual. Wiswesser revised his manual and published a paper in 1953, a book in 1954, and established a Chemical Notation Association to counteract the IUPAC's premature adoption of the Dyson system. But for the aggressiveness of Wiswesser and the work of the Chemical Notation Association, which improved the original system extensively over a period of about 12 years, with the complete rewriting of the system by E. G. Smith in 1968, the Wiswesser linear notation system would have made no impact. The Association with its approximately 75 members continues to develop it and to control and set standards for its use. In 1968, the Institute for Scientific Information began to use the Wiswesser system in some of its products.

Despite the rising interest in the Wiswesser notation system, as evidenced by the relatively large number of publications over the past ten or so years, many chemists have not considered it to be the last word. Thus other notation systems have been introduced, such as by Hayward in 1961, Fugmann in 1961, Skolnik and Clow in 1964, Lefkovitz in 1967, and Skolnik in 1969. Inasmuch as the Chemical Notation Association is married to the Wiswesser System, new notation systems, regardless of their merits, have found very little research or development activity other than that of the innovators. Nevertheless, new notation systems will be introduced in the future that will solve problems that current systems are not able to solve, and hopefully in the near future chemical information scientists will be sufficiently curious and research oriented to evaluate the new against the old.

Nomenclature. Chemistry is an open-end science. From our vantage point of today, we can expect that thousands of new chemicals will be isolated from nature or synthesized in the laboratories for every coming year of the future. At this point in time, over 4 million compounds have been reported and studied in over 20 million references, and, currently, hundreds of thousands are added every year.

Through most of the 19th century, however, as chemistry evolved into a science, chemists isolated, synthesized, and named many thousands of chemicals without an established system of nomenclature. To a large extent, utter chaos was avoided by the authoritative status and influence of a few chemists, such as Berzelius, Liebig, Dumas, Daubeny, Gerhardt, Hofmann, Kekulé, Weltzien, Beilstein, Friedel, and Baeyer. It was not so much that these chemists were in agreement, for actually they were not, but they were in agreement that something had to be done toward systematizing chemical nomenclature. Consequently Kekulé, with support from Carl Weltzien of the Karlsruhe Technische Hochschule, was responsible for having a congress in Karlsruhe in 1860 to discuss the feasibility of a systematic and rational nomenclature with particular emphasis on organic chemistry. About 140 chemists attended, representing most of the European countries, but little was accomplished except to set the stage for a subsequent meeting. This occurred in 1892 in Geneva with highly successful results, viz., the establishment of principles that set the stage for an evolving chemical nomenclature under the aegis of the International Union of Chemistry (IUC), founded in 1922, and later the International Union of Pure and Applied Chemistry (IUPAC). Although a new nomenclature system was promulgated at the Liege IUC meeting in 1930, the extent of the reform was severely limited by the philosophy that "as little change as possible is to be made in terminology universally accepted". Until the publication of IUPAC Definitive Rules in 1957 and 1965, the major authoritative source for chemical nomenclature, especially for chemicals within new classes of compounds, was the office of Austin M. Patterson at Chemical Abstracts Service, and which was known as the Nomenclature Office (now under the direction of Kurt L. Loening).

Nomenclature activity within the American Chemical Society has been expanding appreciably over the past 30 years with practically every division representing a discipline of chemistry, such as organic, inorganic, physical, analytical, fluorine, carbohydrate, and polymer, as well as others, having a nomenclature committee. The American Chemical Society's Nomenclature Committee, a Council standing committee, is the ACS authoritative body through which the divisional nomenclature committees operate before submittal to the NRC Committee on Nomenclature for presentation to IUPAC for final action. The achievement of a uniform nomenclature system appears to be an impossibly attainable goal, yet one that is worthy of our attention and effort and that of chemists for generations to come.

Indexing Systems. Indexing and classification systems are basic to chemical information science. But chemical information scientists over the past 30 years have ignored classification systems, and essentially no innovative work has been reported. This is surprising in view of the great value of classification and taxonomic concepts, such as those of Linnaeus, Mendeleev, and Dumas, to the understanding and teaching of science and to the reduction of many facts to a linguistic matrix of knowledge.

Activity on indexing concepts, on the other hand, has been high, and many innovative systems have been introduced and developed over the past 30 years. There are many kinds of indexes in chemistry, such as subject, author, patent number, formula, and ring. Of these, the subject index and tools for preparing indexes have received the most attention.

Although edge-notched punched and optical coincidence card systems are today a part of our recent past, they played an extremely important role in promoting new indexing concepts, and especially the uniterm indexing system. It was Dr. Mortimer Taube, a Ph.D. philosopher with an M.S. in Library Science, who popularized the uniterm concept in the 1940's and 1950's. Dissatisfaction with library classification and indexing systems and the way documents were filed on library shelves motivated Taube to seek a new approach. Thus, he introduced the uniterm system: the assignment of single or combined words or concepts to represent the contents of a document, the assignment of an accession number to documents, and the posting of the accession numbers on each uniterm card in one of ten vertical columns, numbered 0 through 9, and placing the accession or document number in the column coinciding with its unit digit. Retrieval of references was by inspection of two or more uniterm cards for document numbers in common. Since Dr. Taube described this system before the Division of Chemical Literature in 1951, many papers have been presented and published detailing its application in a variety of ways using index, or optical coincidence tab cards, and finally processing via computer.

Notwithstanding the impact of the uniterm system on chemical information science, particularly the resultant increase of activity in the design of new information systems, it also had some damaging effects. During the 1940's, there was some gratifying work toward making indexing into a science, and some indexes were being designed with a purpose, a logic, and a grammar. This kind of indexing, however, required the indexers to have knowledge of the contents of the documents being indexed, of how information is used, and of what constitutes a good index from an intelligent user's viewpoint.

With the introduction of the uniterm system and the philosophy that the vocabulary of authors contained all the concepts sufficient for retrieval, why indeed have subject specialists trained in indexing prepare the indexes? Anyone who could read was suitable for indexing—all that was needed was to pick out of each document the 20, 30, 40, or more words that an author emphasized and post the terms with the document number for subsequent card or computer retrieval.

Something strange happened as the uniterm system displaced the dictionary card catalog, one of the milestones of the 19th century, which has been the basic tool for indexes during most of this century, the 20th. True enough the uniterm system required many fewer cards than that of the traditional card catalog, yet was elegantly permissive in the number of term cards. But users encountered a new experience: false drops, a term introduced by users of edge-notched punched cards a few years earlier to describe the obtaining of unwanted references in the retrieval stage. Consequently, to counteract the high probability of obtaining false drops, the system was modified with thesaurus control, or what librarians had called subject authority lists, and with links and rolls. Voilà, false drops decreased, but they were not eliminated, and as they decreased toward zero, so did the number of desired references. Thus was introduced two new terms: "relevance" or "precision", the ratio of the number of relevant references retrieved to the total references retrieved; and "recall", the ratio of the number of relevant references retrieved to the number of relevant references in the system.

When information scientists began to use accounting machines, such as the IBM sorter and collator, in the 1950's, then the computer, such as the IBM 9900 and Bendix G-15 in the late 1950's, the IBM 1401 and 7090 in the early 1960's, the IBM 360 in the late 1960's, and the IBM 370 in the 1970's, the uniterm system was the indexing method of choice initially. As the use of computers increased for the processing of information systems, the uniterm system concept was displaced slowly by the keyword concept. At the beginning of the displacement, the conceptual differences were slight—the orientation remained to words with thesaurus control; the major change was at the input and output stages, although the thesaurus began to take on a new look, changing from the conventional subject authority list with definitions to one that listed each word with associated words and denoted as "narrower" (NT) or "broader" than (BT) (same as the conventional "see also"), "use" (same as the conventional "see"), and "related to" (RT) (not equivalent to any conventional term), an RT term implying in many systems that both terms be input.

Inasmuch as the uniterm system reduced appreciably the requirements of the indexer to be a subject specialist, the next logical step was to eliminate the indexer by automatic indexing via computer. This occurred in 1958 when H. P. Luhn keypunched the complete text of an article and processed the tab cards through an IBM 704 to determine word frequency and distribution for the production of an "auto-abstract". In 1959, Luhn reported before the Division of Chemical Literature his KWIC index or keyword in context index, the context being the title and abstract of an article, but generally only the title. Although a certain degree of vocabulary control can be exercised with KWIC indexes, most producers let the chips fall where they will, and we have now almost come full circle to indexes produced with no or little vocabulary control and with no or little intellectual input. Like the skipper who was lost on the Pacific remarking, "with this wind behind us, we sure ought to get somewhere fast", so with the immediate response of online keyword products from the information industry, "we sure ought to get a lot of references fast". But where are we?

Just as Taube rebelled against the restraints of the conventional classification and indexing systems and Luhn against the necessity for intellectual input for abstracting and indexing, some chemical information scientists have rebelled against the uniterm and keyword indexing concepts. Their impact so far has been relatively slight, particularly in the face of the increasing number of on-line products from the information industry based on keyword indexing.

Being critical of uniterm and keyword indexing, but as a long-time user, and after many evaluative studies, I introduced the multiterm indexing concept in 1969 for communicating the informational content of documents by coordination of subject terms and for displaying the generic/specific relationships within certain subject areas, such as catalysts, polymers, properties, and uses or applications of chemicals. The multiterm index has the advantages of the generic nature of classification, the specificity of good subject indexing, the freedom of uniterm indexing, and computer processability. It does not have the input economy of uniterm and keyword indexing in that a subject specialist must assign subject terms for the multiterms. The economic advantages of the multiterm index are in the computer processing and printout costs and, most importantly, in the user's time in finding what he wants without unwanted references.

Possibly because uniterm and keyword indexing was emphasized so much over the past 30 years, some of the more creative chemical information scientists directed their attention and effort to other indexing needs. These efforts yielded the following innovations:

1. Citation indexes. This concept has been practiced successfully in the field of law for relating legal cases to basic precedents, both backward and forward in time. Its importance as a concept for relating chemical references was put to commercial use by the Institute for Scientific Information in its publication *Science Citation Index*.

2. Formula indexes. The inherent weakness of chemical nomenclature for subject indexing prompted the introduction of formula indexes as supplementary tools. Earlier formula indexes, viz., that of Jacobson and Stelzner and that of Hill, cited carbon first, hydrogen second, and other elements next in some specified order. Dyson questioned the advisability of giving precedence to carbon and hydrogen and proposed in 1949 his "Molform Index" which specified the order P, I, F, Cl, Br, S, N, O, C, and H. Fletcher and Dubbs reported on a formula index in 1956 based on the periodic table sequence for the elements except for carbon and hydrogen which they cited last. Skolnik introduced a formula index in 1958 which cited the elements in alphabetical order by their symbols with carbon and hydrogen last. Garfield, using the Hill formula index as input, computer processed it to as many outputs as the number of separate symbols for the production of *Index Chemicus*; he called this the RotaForm Index, publishing a paper on the system in 1963. Another computer produced formula index is HAIC (Hetero-Atom-In-Context) that Chemical Abstracts Service introduced in its 1967 volume index. This index also is computer produced, but only the atoms other than carbon and hydrogen are highlighted in each molecule.

3. Fragmentation Index. Because chemists think in terms of moieties and functional groups for the reactions and reactivities of chemicals, many information systems have been designed with fragmentation codes, arrangements, or classifications as the indexing base. The Beilstein system is an example of one of the earliest of the important methods. During the 1950's and 1960's a variety of methods were introduced for processing tab cards on accounting machines based on assigning a fixed position on the tab card to a specific chemical fragment. An early method for computer processing was based on the permutation of the Wiswesser line notation around a fixed position on the output line. Each of these methods lacked discriminatory power; i.e., none could separate wanted fragments from unwanted ones and none could discern a fragment within a larger fragment. Consequently, I investigated the potentiality of my notation system for the production of a fragment index via computer. Because the notation is written in the same order as structures are drawn and the symbols for carbon include the number of attached hydrogens and the carbon bonds, each fragment, from a methyl group to each specific ring structure, and each functional group are clearly denoted as single units and as related to their neighboring fragments. The notation system furthermore is eminently suitable for the preparation of a notation symbol index that gives considerably more information than does the formula index, the RotaForm index, or the HAIC index, and in addition each molecule is uniquely represented by its notation symbol index in contrast to a formula index which cannot discriminate among molecules containing the same kind and number of atoms.

Computers. As already pointed out, the metamorphosis of information systems through three generations of computers has been relatively easy and logical because the conceptual framework was established during the 1940's and 1950's by systems development with edge-notched and optical coincidence cards. Another positive factor was the ready receptivity to computerized systems by chemical information scientists.

In retrospect, however, there has been a despairing lack of imagination and creativity in the use of computers for processing information. Despite the high activity in the use of computers, most of the emphasis has been on relatively trivial applications and little on the development of new concepts and methods. Progress and advancement in chemical information science are not related to more and faster computers but to greater yield, higher productivity, lower costs, and more reliable and direct answers and results. The computer can be an important tool toward achieving this objective, and will once we understand that it is merely a tool. It is a tool that can do more than manipulate a massive input of information into some ordered sequence. It is one that can be used to provide us with a way of thinking, a way of coping with the unknown, a way of asking questions and interacting with the answers with new questions, and a way of arranging the known so that we may discern the value of an unknown.

The Information Industry. Until relatively recently, publishing, abstracting, and indexing operations and services in chemistry have been associated directly or indirectly with scientific, technical, and trade societies or associations. Over the past 20 to 30 years, however, as these operations and services expanded, many "for-profit" organizations and governmental agencies, universities, and consulting firms have arisen to constitute the information industry. Expansion of the established as well as the introduction of new information services has been supported by the federal government, such as the National Science Foundation.

Thirty years ago the information industry could be thought of as constituting two major segments: producers of primary or secondary publications. With the advent of the computer and government monies, publishers became interested in computer composition of their products and computer processing of the input or a portion of the input for the production of abstracts and indexes. Thus arose the data base producer, followed by the data base processor and broker of products from data processors for on-line users. Over the past few years more than 50 data base brokers have come into existence for on-line services, whose clients conducted approximately 700 000 on-line searches on the data bases in 1974. Examples of these brokers are: the Illinois Institute of Technology, Lockheed, Battelle, Systems Development Corp., and Georgia Institute of Technology. These data brokers supply on-line access to a variety of data bases, such as those of Chemical Abstracts Service, BioSciences Information Service, Engineering Index, National Technical Information Service, and National Library of Medicine.

Data base producers, faced with a mounting literature and inflationary printing and paper costs, strongly recognized the need for cooperation and coordination among each other and with organizations in other nations. An example is Chemical Abstracts Service's success in having the Chemical Society (London) and West Germany's Gesellschaft Deutscher Chemiker assist in the development and operation of a common computerized secondary information system. Such cooperative ventures are but the first step toward an international network. A cooperative program also has been set up whereby three large information services, Chemical Abstracts Service, Engineering Index, and BioSciences Information Service, coordinate coverage and indexing practices. The American Chemical Society, as one of the most important publishers of primary chemical journals and of *Chemical Abstracts*, is now computer processing with one input that information which is common to both, such as titles, authors, abstracts, and registry numbers.

The computer has and will continue to play an important role for information processors and data base brokers. Indeed, we now have more information services and products than ever before. This does not mean, however, that we have greater access to desired information. Unfortunately, information processors and data base brokers have expended government monies on methods and developments, ignoring the basic problems of information storage and retrieval. Government monies have been well spent on the reduction of clerical operations connected with the processing of references for data bases, on the concept of KWIC indexes, and on the development of chemical registry systems. Chemistry, however, is not concerned with references per se but with the relationship between facts, the dependence of facts upon other facts, the systematic arrangement of information, and the underlying principles that govern the relationship between facts. Instead, information services base their systems on the language of the author, pulling from his title the keywords by which the seeker of information can retrieve at most the reference to the document. Publishers and information processors have yet to learn how to contain the chemical literature, to prevent the enormous body of trivial work from diluting the meaningful work, to treat all publications not as equal grist for data bases.

THE FUTURE

I do not harbor the illusion that we can predict the future, for chemistry, chemists, and the chemical literature today are in no way reasonable facsimiles or even projections of a century ago. This is not to deny the past or the present as the force that shapes the future. Remember that the American Documentation Institute was organized in the 1930's by those who anticipated the replacement of records on paper with the microform library. And remember Vannever Bush's 1945 article in the Atlantic Monthly, "As We May Think", in which he stated: "the instruments are at hand which, if properly developed, will give man access to, and command over, the inherited knowledge of the world". We still await the realization of the microform library and on-line access to the world of chemistry. Although we are now pointed in these directions, we are still pretty much at the starting point, with the routes to our tomorrows largely uncharted.

Rather than attempt to predict the future, I suggest that we think in terms of goals. Time does not permit me to do this for every item I discussed in this paper, so I shall restrict myself to a few selected areas which I considered sufficiently important to invite speakers to present papers on.

To many of us, the scientific journal is the milestone of the past that made possible the phenomenal growth of science. It has been the primary channel of communication among chemists of every nation and of every age. I think it will remain so for many decades to come. But the journals of tomorrow will not be like those of today, just as ours of today are not like those of the 19th century. I realize that one of our current educational problems is that reading is becoming a lost art, but I also think that this phenomenon is unique for the television generation of the last 20 years and that it may not mark future generations. I think books and journals, that is, the private library, will always exist to represent each individual's personal choice, tradition, and professionalism. I think our personal contact with scientific books and journals will always be with us for pleasure and self-education, but not to the extent nor in the form they are today. In fact, I find incomprehensible that the world of science supports approximately 40 000 journals, of which about one-third is chemistry oriented or related. Obviously the policies and procedures of our primary and secondary journal structure are not adapted to the world of today, let alone of tomorrow. Our imperfect system, however, will be corrected not by destroying but by improving the structure that carries the burden of our learning. Some day in the future, the journal population will be under control and that which is published will be worth publishing and worth reading. Of course, future generations will have to learn to be better writers than we are and to be willing to submit data to a universal data and information system readily available on-line to every scientist throughout the world. What is published in the journals and read with care will be the meaningful conclusions and paradigms that the data and all related data support.

As long as chemistry remains a rapidly progressing scientific field, we will have nomenclature problems that demand attention. Whatever system of nomenclature we adopt ideally should be open ended so as to accommodate new discoveries without undue upheaval. This is extremely difficult to achieve, and possibly it cannot be. One of our major problems in the past, and also in the present, is our failure to formulate nomenclature for a new and advancing field until after a field is developed. It is not easy to overthrow an existing system of nomenclature. Even though a nomenclature system was rational and logical at one time in the past, but has become less and less relevant as new knowledge unfolds, the tendency is to retain it because it is common usage. Inasmuch as nomenclature is closely associated with our ability to teach and to communicate, a goal of high priority is to design a nomenclature system based on a more natural scheme of classification and relevant to the most current assumptions on the nature of chemical structures.

Closely related to the nomenclature problem is the conceptual framework and applications of notation and topological systems. Although these systems were conceived as answers to nomenclature problems, especially for the indexing of chemicals, they now constitute an area of interest and of importance in their own right. Indeed, some of the best research in chemical information science over the past 15 years has been on notation and topological systems. It is essential that research activity in this area continue well into the future and that chemical information scientists seek new systems not only for retrieving chemicals but also for correlating chemical structures with their properties, reactions, and uses. One of the dangers I perceive is the high desire among chemical information scientists to select prematurely a notation or topological scheme as the answer to all problems. As chemical information scientists, we should know from the history of chemical nomenclature how easy it is for a poor decision to lead to the sacrosanct common usage.

Despite the great importance of classification schemes in the history of knowledge and education, the lack of research on classification concepts over the past 30 years has been a lacuna in chemical information science. This neglect reflects the absence of stress on organized ways of looking at and working with chemistry as a science. Yet science is concerned primarily with the systematic arrangement and organization of facts. Apparently we prefer to relegate to future chemical information scientists the responsibility for designing classification systems relevant to the accumulating facts of chemistry.

Indexing methods and developments, however, have enjoyed a high activity over the past 30 years. Unfortunately, this high activity cannot be viewed as synonymous with progress. Much of the work has been oriented to keyword indexing and concomitant with the increasing number of computer-based information systems. Keywords today are equated to the significant words used by an author in the title of the paper, and sometimes in the abstract or even the whole paper. Abstraction of keywords from the whole text has been a goal of information processors since the introduction of second generation computers, but without economic success. Information processors and brokers currently are enticed with the possibilities of retrieving documents by on-line, whole-text searching for keywords, primarily because this is something computers can do easily. This objective can be economically feasible only if the information processor is tied in with the producer or publisher of documents which are photocomposed via computer. Let us assume that all scientific journals are computer composed suitably for input for whole-text searching by scientists with interactive on-line terminals. We would need to store in the computer system more than one million articles produced annually by about 35 000 journals. The economics is mind-boggling, even for a governmental agency's budget. Moreover, the economics is mind-boggling for the 270 000 journal articles abstracted and indexed by *Chemical Abstracts* in 1973. Most importantly, we need to consider seriously the intellectual consequences of keyword indexing in contrast to subject indexing.

Computers and telecommunications have been a major factor in the changing trends of chemical information science, especially over the past 15 years. Both, in my view, are magnificent tools for the processing and communication of chemical information. They are in no way, however, a substitute for the chemical information scientists' exploration of the fundamental logical nature of chemical information. With the advent of the information processor's and broker's intellectual and financial monopoly of information services, the chemical information scientist's role is changing from the "user of" to the "used by" the computer in the human-machine interface. We may then inquire of the future of chemical information scientists. Let us assume that the word crunching of titles, abstracts, and even the author's complete text in a computer is not a viable hypothesis for chemists seeking information from the literature, that chemists will insist on having retrieval systems that yield only desired references or data, that chemists will require information systems in which they can browse as they do now in a chemistry library, and that chemists will want to interact with the computer as they do now with other chemists. If these are the requirements of chemists vis-à-vis their literature, and I think they are, then chemical information scientists will be with us far into the future working toward the achievement of these goals.

The Journal and Its Possible Future†

B. J. LUBEROFF

CHEMTECH, 48 Maple Street, Summit, New Jersey 07901

Received November 4, 1976

Is it any wonder that in seeking someone to honor for his ability to manage the literature we've picked Herman Skolnik? The man reads it at over 1000 words a minute! I, on the other hand, like most mortals, plod through the literature as though in galoshes. I'm like Julian Huxley's student of foreign language who can't possibly enjoy the "scenery" because he finds himself as though "plodding up a steep hill on a hot day". I was thus more than startled when Herman asked me to address this learned conclave taking as my subject "The Journal and Its Possible Future". It seemed as though an offer to lecture on taxes had been proffered to Crawford Greenwalt who said, in "The Uncommon Man", "I cannot present myself as an authority on taxes except possibly in the same melancholy sense in which a pedestrian is an authority on taxicabs because one has knocked him down".

Nevertheless, I know Herman well enough to know that there is method in his apparent madness. I thus decided that he invited me to share ideas with you not as a pedestrian plodding up the steep hill of the literature but, instead, as an Editor. Why as an Editor? Because they contribute more to

† Presented at the Award Symposium on "Contributions of the Division of Chemical Literature (Information) to the Chemical Society" at the ACS Centennial Meeting, New York, N.Y., April 6, 1976.