more logical primary organization by type of unit process than by product process, which is directly relatable to a unit process. This permits the creation of logical hierarchies and minimizes the sizes of the indexes required for most operations.

PLANNED EXPANSION OF THE ECDIN SYSTEM IN THE CHEMICAL PROCESS AREA

At the present time, two of the three primary files are completed (UNIT/SUBUNIT PROCESS and PROCESS CATALOG files), and the PRODUCT PROCESS file has been filled with data on 76 out of an estimated 382 product processes. Additional work is therefore needed to complete the Product Process file and to develop a user-dialogue display system to permit a user friendly interface to the data. It is expected that outside users will be limited to the use of a set of prestructured queries as they are in the other areas of ECDIN. This permits the system developers to limit access to the actual system. Future efforts are also required to develop an algorithm to permit creation of production/waste trees that will support the display system.

Later, it is envisioned that a PLANT PRODUCT/PROCESS file will be implemented to provide locations for the production facilities employing the product processes by using preestablished map grids. Another anticipated file is that on CONTROL TECHNOLOGY that will be linked to the UNIT/SUBUNIT PROCESS file. The unit process approach is particularly well suited to this logical addition since most control technologies are designed to be applied to sets of similar reactions such as constitute a unit process. It is expected that the CONTROL TECHNOLOGY file will contain data on the efficiency, cost, and manufacturer(s) of the technology and where facilities that use this approach are located.

CONCLUSIONS

The advantages to the user from storing and manipulating process data as developed for implementation on the ADABAS include the ability to consider as a group those chemical processes that operate under similar conditions, the ability to access data on selected chemical processes in a logical fashion, and the ability to combine waste data for an entire production stream. Though the file system described here is only partially implemented, when it is completed, it will provide a powerful accessible, interactive tool for indentifying and solving chemical process waste problems.

REFERENCES AND NOTES

(1) Hushon, J. M.; Powell, J.; Town, W. G. "A Summary of the History and Status of the System Development for the Environmental Chemicals Data and Information Network (ECDIN)". *J. Chem. Inf. Comput. Sci.* **1983**, *23*, 38.
(2) Bletchly, J. D. "A Report on the Apparent Needs of Possible Customers in the United Kingdom for an Operational ECDIN"; EC Joint Research Centre: Ispra, Italy, 1979.
(3) Fourneau, J. P. "Les Utilisateurs Potentiels de l'Environmental Chemicals Data and Information Network (ECDIN)"; Paris, 1979.
(4) Dynamic "PRODUCT PROCESS File"; prepared under Contract 1857-81-05 EP ISP for the EC Joint Research Centre at Ispra, Italy, 1982.
(5) Dynamic "UNIT/SUBUNIT PROCESS File"; prepared under Contract 1857-81-05 EP ISP for the EC Joint Research Centre at Ispra, Italy, 1982.
(6) Herrick, E. C.; King, J. "Catalog of Organic Chemical Industries Unit Processes"; MITRE Corp.: McLean, VA, 1979.
(7) Dynamac "Revised Catalog of Organic Chemical Industries Unit Processes"; prepared under Contract 1857-81-05 EP ISP for the EC Joint Research Centree at Ispra, Italy, 1982.

An Interpretation of Chemical Abstracts Service Indexing Policies†

RUSSELL J. ROWLETT, JR.
Caropines, Myrtle Beach, South Carolina 29577

Received December 5, 1983

Throughout the early years of Chemical Abstracts Service, its document analysts selected index-access points that coincided with where a searcher would be most likely to look first. The growth of the literature, the broadening interests of chemists, and the increased complexity of the science demanded more systematization. Accordingly, six basic indexing philosophies have evolved from the beginning and are still followed today. They are (1) selection of subjects, not words, from original documents, (2) use of molecular formulas as central building blocks, (3) a high degree of specificity, (4) inverted names for organic molecules, (5) continuity, and (6) use of highly trained scientists–analysts for preparation of abstracts and index entries.

Much documentation exists of how Chemical Abstracts Service (CAS) indexes chemical substances and general subjects. The numerous indexes produced over the last 76 years are described thoroughly. The policies that guided index production are well-known and detailed in several publications. I will not refer specifically to any of these excpet as they are needed to understand an illustration.

My goal is to interpret six of the major philosophies that have guided CAS indexing for 66 years and continue to guide this indexing today. An experienced searcher will have encountered these philosophies one by one. I know of no collection and interpretation of their interrelationships.

† 1983 Herman Skolnik Award address, presented before the Division of Chemical Information, 186th National Meeting of the American Chemical Society, Washington, DC, Aug 30, 1983. R.J.R. is a former Editor and Director of Publications and Services of Chemical Abstracts Service.

I said "philosophies that have guided CAS indexing for 66 years". *Chemical Abstracts* (CA) is 76 years old, but indexes as we know them today did not exist during CA's first 9 years, 1907–1916. In the 1952 history of the American Chemical Society,[1] the 43-year Editor of CA, E. J. Crane, wrote, "In a journal as extensive as is *Chemical Abstracts*, information can be buried as far as retrospective searching is concerned unless a really effective index key is provided. With many other developmental problems to handle, the editors apparently did not give special attention to indexing of *Chemical Abstracts* during the first few volumes. Accordingly, when the First Decennial Index was authorized, reindexing by subjects of the first nine volumes was properly undertaken. To provide a model for the First Decennial Index, and for future annual indexes, a special study of subject indexing was made before Volume 10 was indexed. This resulted in a number of changes

0095-2338/84/1624-0152$01.50/0

in policy, and Volume 10 (1916) marks the beginning of emphasis on indexing, especially subject indexing, which more than anything else, has brought recognition of *Chemical Abstracts* as an effectively usable record."

What basic philosophies guided these new policies and this new emphasis on subject indexing? Crane described the innovation as "a new entry-a-line form", which called for "systematic modification writing" to include the significant index words selected from the original document. The most significant selected work was placed first in the index modification.

Many of the basic indexing philosophies that guide CAS today are included in Crane's simple explanation. Some have been adjusted as the science of chemistry and interests of chemists have broadened. Some have been expanded to accommodate modern technology in information transfer. However, the basics are still with us.

CAS indexes subjects, not words selected from the title or the document. Index entries evolve from the body of the original document. These workds may or may not be contained in the abstract. Thus, CAS indexes the significant content of reported research and technology not just the brief contents of abstracts. Usually, the collection of all index entries from a document represents a more complete summary of the contents than the abstract alone. This is especially true for documents that describe chemical syntheses. In such cases, hundreds of substances are indexed, but the abstract contains only a few, with illustrative formulas.

This philosophy of indexing subjects from the original, even though the subjects are omitted from the abstract, has raised many questions over the years, particularly in organic chemistry. Editor Crane felt strongly that all new substances for which new information was given should be recorded in the abstracts. This decision led to very lengthy abstracts of organic chemistry. The result pleased few readers. Early in CA's history (1912), the American Chemical Society Council actually voted that organic abstracts were too long. Today, all new substances are indexed. All old substances are also indexed if there is new information reported about them. New information takes many forms, for example, new sources or methods of preparation, new chemical, physical, or thermodynamic properties, new reactions with the substance, new reaction kinetics and mechanism studies, new uses and applications, and new health and safety information.

New information about a substance has sometimes been referred to as "novelty". Also, it is said that a primary criterion for indexing is "novelty". General subjects are indexed when they reflect or highlight the novelty of data contained in the author's report. However, in all disclosures of new information, the novelty claimed must be substantiated by some documented data. The latter takes many forms and is not just numerical or physical measurements. It is often a discussion of reaction mechanisms or the probable structure of the molecule being reported. For CAS to index a substance or a subject, there must be some hard information present other than just a name or a structure. If a searcher consults that reference, something novel about the substance or subject will be found.

Chemistry is a unique science. There is an international language useful for exchange of information and as a hub around which indexes may be assembled. That language is the structural molecular formula. CAS indexing was built around those formulas and their corresponding systematic names. The development of the computerized Registry System made this early decision of even greater significance. Over six million substances, identified since 1965 in regular CAS indexing operations, can now be searched on-line in a fraction of the time once required to review a single volume index.

No other science has such a hub around which to build its information system. Perhaps this is the main reason why no other scientific information systems are as effective as are those for chemistry. As other sciences have become more molecularly oriented, scientists in fields such as physics and biology have turned often to searches of the molecular-substance-oriented data base of CAS.

Another philosophy inherent in Crane's early statement is that CAS indexes include a high degree of specificity. All substances and subjects are indexed as specifically as possible, including stereochemical details. The systematic index names are rigidly controlled. Additionally, general-subject index terms are controlled as tightly as is possible considering the rapid growth of the science. An individual document analyst cannot enter a new index heading without it being reviewed by a second subject expert in that particular field. New substance formulas and names are subjected to a series of computer edits, any one of which may reject the name and formula for additional human review. Author-used names for substances or properties are cross-referenced to controlled index terminology.

This specificity in indexing as it relates to use of rigidly controlled vocabulary has not been true for the entire life of CAS. As late as the 1950s, CAS attempted to index a substance or a subject at the index location where the searcher was expected to look first. While this may have been satisfactory for annual indexes of a few thousand pages, it resulted in far too much scatter of similar information in the huge, multivolume, annual indexes that evolved in the 1960s. Today, the actual index production would not be possible without strict adherence to the selectivity philosophy backed by the rigidly controlled vocabulary and its system of cross-references.

The desire to place references at index headings that a user would most likely consult first led ultimately to the use of thousands of trivial (nonsystematic) names for chemical substances. For a complex molecule, it was often possible to derive several completely descriptive names on the basis of different trivially named substances. No longer could editors be sure where the searcher might look first. Nor could editors be sure which name was most descriptive of the molecule's chemical activity. The result was a major revision of CAS substance index names in 1972 as the Ninth Collective Index period began. Most trivial names were dropped, and fully systematic nomenclature was followed wherever such nomenclature principles existed.

An interpretation of CAS nomenclature philosophies is a subject too large for discussion here. However, two early nomenclature-related decisions continue to have great influence on basic indexing philosophy. Austin M. Patterson made contributions to both decisions. In 1916, with CA Volume 10 and with the subsequent First Decennial Index, Patterson instituted the use of inverted index names for organic substances. This technique spawned what we know today as "CAS index heading parents". These are simply the uninverted part of Patterson's index names. Their use resulted in the assembly of information about similar substances at nearby locations in a printed index. Today, structural representations of the univerted parts of the names help make possible computer searches that Patterson and other early fathers never dreamed of.

The use of inverted names demanded selection principles for what is today described as the heading parent. Thus arose the CAS name selection principles. Again, Patterson had a leading hand. He surveyed a segmeent of the organic chemical literature and determined how most chemists named a selected group of chemical molecules. He thus answered the question of where the average searcher might look first for a particular molecule. Patterson repeated this survey for the first four

Decennial Indexes, reviewing and revising the list according to the name usage that he found in the original literature. His "order of precedence of functions" is basic to todays CAS name selection principles. The latter has been revised over the years for new molecular systems and the needs of the expanding science of chemistry.

Index entries are made for classes of substances when the document reports several members of the class or when the class is the object of the study. When to make a class index entry is a difficult question for the document analyst to answer. Some users want a class entry for every substance and subject, particularly for those that pertain to their individual special interests. This, of course, is neither practical nor financially viable. Such an effort would comprise what has been called "bibliography indexing". It would maintain all references to a few users' favorite subjects. CAS has rarely been trapped into such indexing. The worst example is the old index heading "Spectra". A former Editor was talked into making an index entry at Spectra every time the then new research tool was used and reported. The ultimate result was a huge heading, unmanageable and unuseful. The practice was abandoned.

Experience with the Spectra index heading illustrates two facts about CAS indexing. At the time the Editor agreed to make the Spectra heading absolutely complete, there was only one type of visible spectra and few actual entries. The decision appeared to be useful to chemists generally. However, rapid expansion of use of the new tool plus development of many other types of spectroscopy rendered the decision worthless. The growth of the science of chemistry has a demanding influence on indexing and must be watched constantly. Today, in CAS Volume Indexes, 57 different index headings are needed to cover the interests of scientists in the still growing field of molecular spectroscopy. These 57 headings are another example of the specificity employed by CAS in indexing a complex subject. Obviously, the old general index heading would be even less useful today than it was 30 years ago. Similar examples can be described from many other chemical subjects.

Throughout its history, CAS has emphasized continuity of indexing. This is the ability to go backward in time through the annual or collective indexes and be able to find references to the same substance or subject. This has not been easy because of rapid expansion of science and the continued year-after-year synthesis of over 350 000 new substances. Corss-references are sign posts that guide the searcher in the index continuity. Since 1968, cross-references have been contained in the CAS Index Guide. Prior to 1968, they were included, and repeated, in each and every index volume. User demands have recently led to reinstatement of a few important cross-references in the Volume Indexes.

Continuity is present. It has been maintained very carefully. However, some feel the "mapping" has not been as thorough and as clear as it might be. This is an effort that can be improved greatly in the CAS ONLINE computer data base, particularly as the files prior to 1965 are input for substance registration and, hopefully, later as the corresponding index and abstract contents are input. Mapping backward in time with a science that is expanding rapidly in subject as well as growing in total volume is not an easy task.

Five basic CAS indexing philosophies have been discussed: (1) selection of subjects, not words, from original documents; (2) molecular formulas as the central building blocks; (3) a high degree of specificity; (4) inverted names for organic molecules; (5) continuity. There are others, but these are five of the most important. However, none of these five would have been possible without a sixth, use of highly trained scientists–analysts for preparation of abstracts and index entries. Document Analysts is their current title. In former years, abstracts were prepared by several thousand volunteers around the world. They were subject and language experts in their particular fields. Indexers did their job after abstracts had been printed. Today, a single subject expert reads, understands the original document, and then prepares the abstract and index entries in a single intellectual step. Automated computer edit and support systems assist and make the single step possible. A better index results, and it is distributed more quickly, ready for use in a growing number of output forms. The selection processes for both documents to be covered and preparation of abstracts and index entries continue to be in the hands of thoroughly trained scientists. I expect this will continue as the computer grows in its ability to relieve scientists of most mechanical chores. Scientists can then concentrate more fully on the intellectual efforts of interpreting and recording the author's work for posterity.

When CA began, only humans were available for such tasks. They will not be replaced by even the smartest machines. Smart machines will assist smart scientists to perform CAS indexing more thoroughly, more efficiently, and more rapidly.

I can not really foresee the whole process becoming less costly. I have been told recently that the cost of human indexing is too great and we must move entirely to machine indexing. I do not believe this. I know without human control of the selection process the index quality will deteriorate rapidly. This is especially true in chemistry where new substances have to be understood, interpreted, and entered into the data base. There are author errors and misunderstandings in the original documents that appear in well over 50 languages. The continued use of trained scientists for document analysis is my sixth basic philosophy of CAS indexing.

REFERENCES AND NOTES

(1) Crane, E. J. "A History of Chemical Abstracts, A History of the American Chemical Society"; Chemical Abstracts Service: Columbus, OH, 1952; p 19.