A History of General Subject Indexing at Chemical Abstracts Service

D. F. ZAYE, W. V. METANOMSKI,* and A. J. BEACH
Chemical Abstracts Service, Columbus, Ohio 43210

Received December 3, 1984

General subject indexing for *Chemical Abstracts* began in 1907 and has been expanded, improved, and refined over the years. These changes occurred incrementally to reflect the growing complexity of chemical methodology. The development of various indexing policies has been guided by the Editors of *Chemical Abstracts* with each leaving his own mark on the structure and organization of the indexes. The indexing and editing operations have progressed from a purely manual effort to one that is today highly automated. Chemical Abstracts Service has endeavored to be responsive to the needs and desires of index users. This has prompted the development of detailed and comprehensive instructions on the use of the indexes, as well as workshops and workbooks, and presentations and papers on indexing in specific areas of chemistry and chemical engineering.

INTRODUCTION

Chemical Abstracts (*CA*) began publication in 1907 under the editorship of William A. Noyes, Sr. The history of abstracting at Chemical Abstracts Service (CAS) has been traced in a publication by Baker et al.[1] This paper tracks the development of general subject indexing. CAS produces many kinds of indexes varying in both content and format. Because of this diversity, only the indexing and editing procedures for general subjects are addressed here with focus on specific characteristics of the general subject indexes and how they are produced. Chemical substance indexing, an extremely significant undertaking with respect to CAS coverage and the effort devoted to it, will be covered in a subsequent paper. Figure 1 defines and cites examples of general subject headings, occasionally referred to as concept headings.

THE EARLY INDEXES

In December of the first year *CA* was produced, a volume *Subject Index* and accompanying *Author Index* were produced for the 24 issues. This *Subject Index* was similar in appearance to, though much less sophisticated than, the current indexes. Prior to 1972, volume indexes combined chemical substances, general subjects, and cross-references into a single listing. Since then, the Subject Indexes, because of their huge size, have been divided into these three components. The number of index entries for Volume 1 (1907), almost 19 000, represents a mere six-tenths percent of the 1983 count.

Of the 7850 index headings in Volume 1, 75% contained a single reference. The large majority of these singleton entries were chemical substance names. The remaining 2000 index headings, some of which had over 140 postings, had no particular sort order to the index modifications. Some were ordered alphabetically, others by *CA* page number, and still others by the sequence number of a series of published studies. Users were expected to scan all modifications at a given heading. This was not as time consuming as it might appear as there were, on average, only six or seven entries per heading, discounting those headings that contained only a single reference. As seen in Figure 2, the index entries always have been "articulated". That is, a simple rearrangement of the words in the index entry will produce a phrase, corresponding to a near natural language sentence, that summarizes one aspect of the referenced study. Articulated indexes are more time consuming to produce, but each entry conveys the needed information in a context that is not possible in keyword indexes, which lack syntax.

From the very first volumes, index headings for general subjects have been controlled, and cross-references directed users from synonyms to the index headings. Some of the headings used in the first indexes are still presently in use. Also, many formatting practices were initiated in the earliest indexes that have proved to be useful even in the more highly structured indexes of today. Index headings several columns in length were subdivided to make searching faster and easier. For example, at the index heading **Apparatus**, in Volume 3 (1909), subheadings, such as "electrical", "filtering", and "photometric", among others, were found. Inverted headings were used in some cases, such as **Conductivity, electric** and **Conductivity, thermal**, to keep similar concepts near each other in an alphabetically sorted index. Homonyms were not identified as such.

One striking difference between the first index and subsequent ones was in the choices of index headings. In a few instances, examples of word indexing, rather than subject indexing, could be found. Word indexing is valuable in building a concordance but can lead to problems for index users. Subject indexing consists of the indexing of the subject reported by the authors but does not necessarily use their words. This procedure was introduced in the *First Collective Index*, described below, and continues to this day.

The development of indexing policies, especially for chemical compounds, was begun by Austin M. Patterson, *CA* editor from 1909 to 1913. Under the leadership of E. J. Crane, who became the *CA* editor in 1915, a solid foundation for the future of *CA* indexing was laid. He recognized that greater attention had to be paid to indexing practices to produce a "thorough and thoroughly good" index.[2] Hence, in 1916, he directed the preparation of the *First Decennial Index*. Prior to this, *Subject Indexes* had been provided annually, but these were limited in scope, author terminology was used that resulted in the scattering of information, and formatting of the index entries was not consistent.

To begin the effort of building a collective index, a special examination was conducted to develop a model. On the basis of this, it became necessary to reindex the nine volumes that had been issued previously. To justify this tremendous amount of work, it was necessary for Crane to present his case to the American Chemical Society. The Society granted its approval after he had gathered almost 2000 advanced subscriptions for the proposed collective index, at $10 each. Crane himself then did much of this reindexing. Because production of each of the succeeding collective indexes has been costly, subscriptions have continued to be solicited in advance.

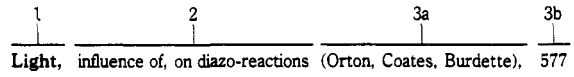
Subject indexing, as contrasted to work indexing, was adopted during the preparation of this first cumulative index. This resulted in fewer omissions of relevant entries, bringing together similar kinds of information, and a decrease in unneeded entries. Related studies were also indexed uniformly. Crane's constant emphasis on consistency[3-6] led to indexes that, despite the difficulty of the subject matter, were relatively easy to use. Beginning with Volume 10 (1916), an introduction

0095-2338/85/1625-0392$01.50/0

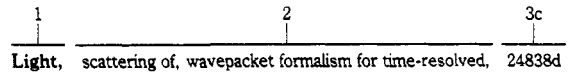
Types of General Subjects	Examples of Index Headings	
Animals	Dog	Salmo gairdneri
Apparatus	Condensers	Mixing apparatus
Applications	Silvering	Staining
Chemical groups	Formyl group	Phosphate group
Chemical substances (non-specific)	Amber	Linseed oil
Classes of chemical substances	Alkali metal halides	Glycols
Classes of property- or use-oriented materials	Appetite depressants	Surfactants
Diseases	Epilepsy	Silicosis
Effects	De Haas-van Alphen effect	Stark effect
Laws	Boyle's law	Raoult's law
Materials	Bituminous materials	Building materials
Microorganisms	Bacillus albidus	Escherichia coli
Organs	Brain	Kidney
Plants	Canavalia ensiformis	Gooseberry
Processes	Extraction	Handling of materials
Properties	Electric resistance	Viscosity
Reactions	Grignard reaction	Oxidation
Rocks	Diabase	Granite
Scientific disciplines	Astrophysics	Geochemistry
Miscellaneous	Awards	Biography

Figure 1. Examples of *CA* general subject (concept) index headings.

Index entry from Volume 1 (1907) *CA* Subject Index:



Index entry from Volume 98 (1983) *CA* General Subject Index:



Key: 1. Index heading (controlled vocabulary)
2. Modification (free text)
3. a. Reference (author names)
b. Reference (*CA* page number)
c. Reference (*CA* abstract number)

Figure 2. Examples of index entry formats from past and present indexes.

to the index was also provided. The introductory material to the indexes is still being provided today, revised and updated as necessary. Volume 10 users also saw the beginning of the "entry-a-line" format; that is, each index entry was started on a new line. This format continues to be used today.

GROWTH OF THE INDEXES

The size of the indexes continued to grow, reflecting the increase in the number of abstracts in *CA*.[1,7] The *Subject Index* for Volume 1 (1907), 201 pages, nearly tripled in size by 1911 but declined to 277 pages in 1916 (World War I period), at which time the *First Collective Index* was compiled. *Collective Subject Indexes* continued to be compiled once every 10 years until 1956 when the collective period was cut in half. The sustained growth of the indexes prompted additional measures to promote manageability. In 1972, at the onset of the ninth collective period, the *Subject Index* was divided into two parts: one dealing specifically with general subjects and the other with chemical substances. A separate part, the *CA Index Guide*, contained cross-references and indexing notes. The publication record of the *CA* Collective Indexes is depicted

Table I. *CA* Collective Indexes

	Subject Index		*Index Guide*		*Author Index*		*Formula Index*		*Patent Index*		
CA collective period	no. of books	no. of pages	no. of books	no. of pages	no. of books	no. of pages	no. of books	no. of pages	no. of books	no. of pages	total books
1st (1907–1916)	2	2843			2	1980					4
2nd (1917–1926)	3	4139			2	2452					5
3rd (1927–1936)[a]	3	4885			2	3095					5
4th (1937–1946)	4	6386			2	3541	2[c]	2077	1	182	9
5th (1947–1956)	11	13740			4	5074	3	2968	1	144	19
6th (1957–1961)[b]	7	12659			4	6164	3	3869	1	172	15
7th (1962–1966)	13	24632			6	9607	4	7007	1	352	24
8th (1967–1971)	18	33548	2	2280	8	13845	7	11769	*d*	761	35

	General Subject Index		*Chemical Substance Index*		*Index Guide*		*Author Index*		*Formula Index*		*Patent Index*		
	no. of books	no. of pages	no. of books	no. of pages	no. of books	no. of pages	no. of books	no. of pages	no. of books	no. of pages	no. of books	no. of pages	total books
9th (1972–1976)	10	16797	25	40891	1	1140[e]	11	19128	9	16251	1	1374	57
10th (1977–1981)	15	25508	32	56112	1	1605	14	24684	11	19904	2	3643	75

[a] Changed into a larger format and three-column page. [b] Changed into a larger format. [c] Covering 1920–1946. [d] Combined with *Author Index*. [e] Illustrative structural diagrams, published in the *Parent Compound Handbook*, were removed.

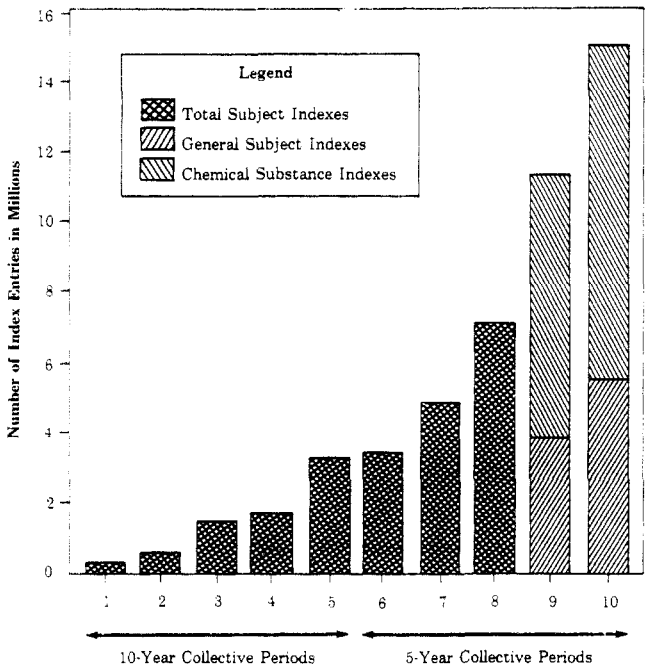


Figure 3. Number of general subject and chemical substance index entries in *CA* collective indexes: first through tenth (1907–1981).

in Figure 3 and summarized in Table I.

For many years *Subject Index* entries had been drawn from the corresponding abstracts. This meant that abstracts had to be as complete as possible to ensure sufficiently thorough indexing. Crane repeatedly emphasized this completeness:[8]

> "Since a complete and permanent record is the object [of the ideal chemical abstract] it is very important that the abstract should contain or make specific reference to all of the information in the paper that is suitable to be indexed. This would include every measurement, observation, method, apparatus, suggestion and theory which is presented as new and of value in itself."

And[9]

> "The policy of completeness is reflected in the requirement that abstracts be written so as to be complete from the indexing point of view and in the special emphasis which is placed on thorough indexing."

Further,[10]

> "Because of the belief that a record such as an abstract journal constitutes is of greatly lessened value if the information is not made readily and certainly available by thorough indexing, more attention has been given to the indexing, particularly the subject indexing, of *Chemical Abstracts* than to any other feature."

Complete abstracts, while always the goal, could not always be prepared. There were, at times, financial problems that imposed limits on the length of the abstracts, and such uncontrolled factors as paper shortages and wartime disruptions. To compensate during these exigencies and to hold lost information to a minimum, indexers were encouraged to make indexes as complete as possible.

Beginning in 1970, CAS adopted a policy that ensured that all abstracts were uniform. Thus, in some *CA* sections the abstracts, which had been very long and detailed, were shortened. This did not affect the depth of indexing, however. Index entries were drawn from the original documents whenever possible and in many cases had no direct equivalent in the corresponding abstract.

INDEXING POLICY DEVELOPMENT

Because of Crane's high standards for quality indexing, upheld by Charles L. Bernier, *CA* Editor from 1958 to 1961, the development of numerous rules was unavoidable and actually encouraged. They felt that standardization was[5]

> "achieved in a subject index built by rule by recording each decision that affects the index. In the *Chemical Abstracts* office if the note or rule is of general interest, it is replicated and given to all indexers; if not, it is used by the indexer making the rule to increase his precision of indexing."

Crane drew up a list of seven general rules for formulating index entries. For example, only new information, or information reorganized for use as comprehensive reviews, was to be indexed. Index entries were to be made to maximum specificity, and new index headings were created only after careful thought. These rules are still in effect today.

However, to be successful as an indexer and editor at CAS meant that hundreds of rules had to be learned and many committed to memory. The activity contributing the most rules was the writing of index modifications, that part of the *Subject Index* entry following the heading and explaining the context of the heading in the referenced study. Here Crane drew up six general rules. Some were very straightforward: "Alphabetical order is used wherever possible" and "Unnecessary symbols and terms are omitted, especially at the beginnings of modifications." Others appeared arcane: "When two or more terms in conjuction are associated on an equal basis with a third term, the modification is started with the third term even though this violates rule 1 [on placing the most significant word first]." But, they were effective given the situations for which they were devised.

Index heading changes are made only after careful examination of literature usage, existing indexing procedures, and, whenever possible, discussions with "bench" chemists. When changes are needed for index headings—to update terminology, to be more exact, or to place entries where users were most likely to look—they are made at the beginning of the collective period. An example is illustrated in Figure 4 for the *General Subject Index* heading **Chromatography**. Although the term was first coined in 1731[11] and used by Michael Tswett as early as 1906 in his paper on the separation of pigments,[12] it was not until the 1930s and 1940s that the technique was extensively developed. In the *Third Decennial Index* there were 16 entries at **Adsorption**, "chromatographic"; 13 entries were at **Analysis**, "chromatographic", as well as a cross-reference from **Chromatography** to those two headings. In the *Fourth Decennial Index*, **Chromatography** became a heading in its own right with over three columns of postings. In that index, the cross-references went in the other direction: from **Adsorption**, "chromatographic" and **Analysis**, "chromatographic" to **Chromatography**.

During the next collective index period from 1947 to 1956, the heading was renamed **Chromatography and Adsorption analysis**, and it contained over 19 pages. Most studies were either improvements in the method or its use for the separation of different substances. In the latter case, the particular kind of chromatography (i.e., gas, paper, column, etc.) was not identified. Additional cross-references were used to reflect the growth and application of the technique. Among these were **Anions**, "chromatography of", **Cations**, "chromatography of", and **Separation**, "by chromatography". But because of the burgeoning size of the heading, two strategies were used to make it more manageable to searchers, and these were implemented during the next two collective index periods. In the *Sixth Collective Index* the scope of the heading was narrowed to include only studies on the technique itself. The use of chromatography simply as a means of separation of two or more substances did not warrant an entry at the **Chromatography** heading. A note appeared at the heading telling users of this:

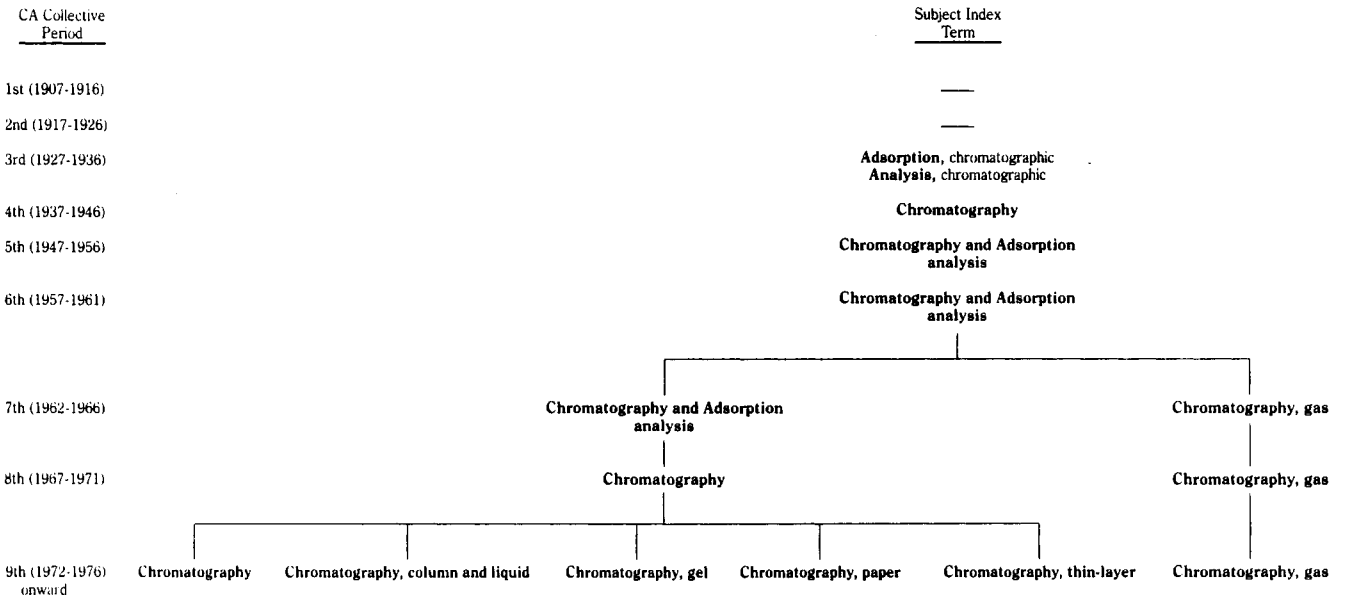


Figure 4. Growth and segmenting of a subject index heading.

> "New techniques, including those on anions and cations, and relationships of chromatography are entered here. See also headings for the specific substances and classes of substances that have been treated by this technique."

The second strategy adopted during the next collective index period was isolating and separating entries on gas chromatography and moving them to a separate heading. This arrangement was satisfactory until the ninth collective period, beginning in 1972, when once again it was necessary to subdivide the **Chromatography** heading into five narrower index headings, while retaining the **Chromatography** heading for very general or undefined studies. These two methods, redefining the index heading and segmenting it into smaller, more specific headings, are commonly used. It must be remembered that it is necessary to correlate cross-references whenever changes of these kinds are made. For chromatography studies, the number of cross-references has increased from 2 (*Fourth Decennial Index*) to over 60 (*Tenth Collective Index*).

CAS follows the policy of selecting the most specific index heading possible for each given subject, be it a reaction, property, or apparatus. For instance, the *General Subject Index* headings **Alkylation**, **Magnetic susceptibility**, and **Absorption apparatus** were introduced in the *First Collective Index* and are still valid today.

From the earliest indexes, guidelines for constructing index entries have been prepared for editors at CAS. When the staff was few in number, changes were communicated orally. As the number of editors increased and responsibilities varied, word-of-mouth was neither efficient nor effective. During the 1940s, office memoranda, detailing rules for indexing and editing, were distributed as needed to staff members. These guidelines were never distributed externally, however, with the exception of the rules for naming chemical substances. Most of the memoranda dealt with routine technical questions, such as index entry placement and format as well as term usage and nomenclature procedures.

Much thought was also given to the creation of new headings and the types of studies to be indexed at those headings. However, Crane resisted the idea of compiling a comprehensive list of all index headings. According to him,[5]

> "To a large extent it is possible to derive subject-index headings by rule rather than by list. This is done in subject-indexing *Chemical Abstracts*, for example. Use of rules saves the indexers' time needed to consult a list, and avoids the obsolescence and cost of lists."

and[4]

> "In the building of each annual subject index *Chemical Abstracts* regards the index of the preceding year as a preferred substitute for a subject heading list. In indexing the whole of chemical progress annually we need room for growth and flexibility. ...A subject heading list built up by one group of indexers would be useful to others only in case the other indexers wish to cover the same kind of subject matter for the same general purpose."

Nevertheless, as computer technology came into widespread use such lists were easily prepared and found to be very useful in indexing and editing. And arranging the list of index headings in hierarchical order proved to be a valuable search aid. Users could see how index headings were related, and since indexing is done to maximum specificity, searchers could select the appropriate, narrowest heading for locating studies of interest.

Computer-produced lists also proved beneficial to CAS document analysts by bringing together those groups of *General Subject Index* headings that receive special treatment. For example, definitions are required for homonymous index headings to avoid errors in placement. By listing together all homonyms along with their respective definitions, analysts can determine immediately the correct heading assignment from the controlled list.

Another indexing concern is those headings that had an extremely large number of postings but that did not lend themselves to being subdivided into a series of new but related headings, as was done with **Chromatography**, mentioned above. The problem of handling these large headings, as well as other indexing problems, led to the selection of a consultant, Russell J. Rowlett, Jr., in 1965, who extensively studied *CA* index structure. He recommended, starting with the eighth collective period in 1967, that these headings be subdivided according to the context of the study reflected in the index modification. Examples of these and other types of subdivisions used for different groupings of headings are illustrated in Figure 5.

During the 1950s and 1960s as CAS increased its staff in response to the growth of the chemical literature, the number of memoranda covering technical procedures and processing operations grew correspondingly. Even after codifying and indexing the memoranda, a more systematic way was needed to sort and categorize this broad spectrum of information. In the mid-1970s this collection of memoranda was replaced by the *Editorial Analysis Manual*, which brought together in one place information previously scattered among several files. The

Index Heading	Subdivisions	
Headings with subdivisions* based on context of study		
Acids		
Clays	analysis	properties
Organic compounds	biological studies	reactions
Polyester fibers	occurrence	uses and miscellaneous
Salts	preparation	
Transition metals		
Headings for classes of compounds with subdivisions based on functional groups		
Fatty acids	anhydrides; esters; hydrazides	
Glycols	esters	
Headings for alloys with subdivisions* based on composition		
Alkali metal alloys		
Rare earth alloys	base	
Semimetal alloys	nonbase	
Headings with biological organ subdivisions*		
Brain	composition	neoplasm
Kidney	disease or disorder	toxic chemical and physical damage
Liver	metabolism	
Headings with subdivisions* based on effects		
Electric discharge		
Gamma ray	biological effects	
Laser radiation	chemical and physical effects	
Plasma		
Sound and Ultrasound		
Headings requiring homographic definitions		
Bombs	apparatus; explosives; volcanic	
Mole	animal; gram molecular weight; neoplasm	

*Each subdivision can apply to each heading.

Figure 5. Examples of subdivided general subject index headings.

present version of the *Editorial Analysis Manual* covers such diverse topics as the technical content of *CA*, the mechanics of abstracting, indexing, and editing, and the details of the computer-support systems.

INDEXING AND EDITING OPERATIONS

The earliest records detailing CAS processing operations have long since disappeared. It is very likely, however, that processing was almost unchanged from the first index produced in 1907 until the mid-1960s when CAS began full-scale automation. Up to that time, all processing was done manually.

Bernier and Crane described the operations at CAS during the late 1940s and these operations very likely were used for the previous decades:[3]

> "For many of the entries in the *Chemical Abstracts* Subject Indexes, eight operations by eight different people are involved...(1) marking page proof for indexing, (2) making cards, (3) checking cards for references, (4) checking these for modification writing and subjects selected, (5) alphabetizing the cards, (6) editing the assembled cards, (7) checking the references on proof, and (8) general proofreading."

Each of the eight discrete steps mentioned above does not appear particularly time consuming when an individual index entry is considered or even a set of entries derived from a given study. However, the time needed to accomplish the steps taken together quickly adds up.

Prior to 1947, index entries were written or typed on 3 × 5 in. cards, one entry per card. The use of wire recorders was begun in 1947, the dictated entries being transcribed clerically onto cards for editing. Again, Bernier and Crane:[3]

> "The editing of the assembled cards at the end of an index period is much more of a task than might be expected. A group of alphabetized cards under one heading presents a different picture from that presented by a single card when it is made or checked. This editing for *Chemical Abstracts* is aided by an extensive system of inverted cross references (40,000 of them) to ensure that the correct cross references are used, to get like things together, and otherwise to provide careful correlation."

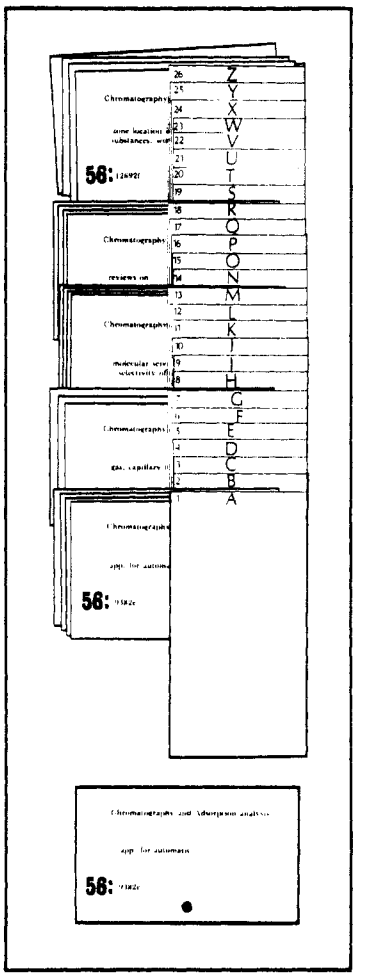


Figure 6. Alphabetizing board with 3 × 5 index entry cards.

To speed up this enormous task, "alphabetizing boards" were used (Figure 6). The cards containing the index entries were color coded for each *CA* volume inasmuch as index entries from more than one *CA* volume would be edited simultaneously. The number of different volumes being edited at the same time was dependent upon the time of the year as well as the year within a collective period.

The editing of large index headings was time consuming and tedious, but singleton index headings presented a different set of problems. If an index heading contained only a single entry, several previous volume indexes were checked to make sure the heading was valid. As mentioned previously, no authority lists existed that editors could use to validate these singletons. If the index heading was not found in one of the previous indexes, the entry was pulled from the editing process and sent to a specialist who would determine whether the introduction of a new heading was justified. Presently, index headings that do not match those on a computerized authority file are sent to a specialist for resolution. All others continue by normal processing routes, thus resulting in tremendous time saving as well as improvement in the overall quality of the volume and collective indexes.

Singleton entries commonly occurred in the indexing of taxonomic terms. Of the approximately 62 400 *General Subject Index* headings presently in the authority file, 54 700 are taxonomic. Looking at this situation from a slightly different perspective and to show how much these index headings, which generally contain a single entry, affect editing, of the roughly 625 000 general subject index entries presently in a semiannual *CA* volume, only 5% are taxonomic. Since 1972, the staff of the BioSciences Information Service (BIOSIS) has provided invaluable help in validating taxonomic

terms new to the CAS file as one aspect of cooperative agreements between the two organizations.

The editing of index entries for a whole volume of *CA* was a complex task and required many separate steps. Editors would use the sorted index cards and the corresponding Master File items—editing notes, cross-references, and scope notes—for that set of index headings. The editing notes indicated the kinds of studies to be indexed at that heading and general formatting considerations: important words that were to appear at the beginning of index modifications, appropriate indenture phrases, and so on. Index entries at a given heading were examined to see if any specific terms were present that would justify the insertion of a cross-reference. Scope notes, which described the kinds of studies the index heading encompassed or explained what users should take for granted when examining a given index heading, were merged with the set of entries, all of which were then forwarded to the printers.

Collective indexes were produced for 10-year intervals until 1956 and for 5-year intervals thereafter. The cumulative indexes were[15]

> "not just a compilation of annual and semiannual indexes. [They] correlate developments which have occurred in the interim and incorporate important changes brought about by the standardization of nomenclature, the introduction of new terms and the crystallization and refinement of others."

Crane further described this process of building a collective index, pointing out that[13]

> "*Chemical Abstracts* reports new chemical information. Much of this information has to do with pioneering effort in fields in which the nomenclature is naturally undeveloped and unstandardized because of the newness of the subject matter and the lack at the early stage of information needed for nomenclature standardization. This means that the indexer is confronted with many nomenclature problems in building a collective index, problems which are not limited to the naming of compounds. The nomenclature of all branches of science in which chemistry plays a part must be considered."

According to Melvin G. Mellon,[16] "For effective use the semiannual and collective indexes are the most important part of *Chemical Abstracts*."

Thus, as subject areas became more understood and more widely used, chemical terminology would be adapted as necessary. This upgrading of material as it was assembled for the collective index could magnify small problems or create new ones that would not exist in the individual volume indexes. This was a result of the very large body of information that was being examined and processed. In an article entitled "Why Indexers Turn Gray", Crane took a light-hearted approach in surveying these issues.[14] However, editing in context made it possible for one editor to view all the entries under one heading (and its neighbors) and to collect similar index modifications at the same place. This, of course, eliminated scattering of references to related studies. During the eighth collective period (1967–1971), computerization was introduced, and this last labor-intensive step was eliminated.

COMPUTER ASSISTANCE IN INDEXING

The first stage in adding machine assistance to the indexing procedure was the final editing step, described above. The groundwork for computer applications was laid in 1955 when CAS formed a Research and Development Department, the first organization of its kind to do so. The Research and Development Department was initially directed by Karl F. Heumann (1955–1959) and then by G. Malcolm Dyson (1959–1963). Under Dale B. Baker, who succeeded Crane as CAS Director in 1958, production operations of *CA* have undergone many changes to take advantage of technological developments and thus maintain currency of the journal and accuracy of its content:[17]

> "It became evident that the traditional, manual system for processing and publishing secondary chemical and chemical engineering information was too slow, too expensive, too rigid, and too wasteful in its use of highly trained manpower to be effective in the face of the ever growing volume of published information. A streamlined operation was needed in which each primary document would be analyzed only once and the selected data recorded only once in a form that could be used to produce a variety of information packages with overlapping content. We also wanted a system that could easily be modified to fit changing needs and one that could readily produce additional publications and services. We defined a computer-based processing system that meets these basic requirements and...began implementing this system in a careful, stepwise fashion so as not to interrupt the continuity of existing services."

Improvements in the operations dealing with the composing of the indexes begain in the early 1960s. Monotype composition, done by Mack Printing Co. of Easton, PA, was replaced by cold-type composition at CAS. During the seventh collective period (1962–1966), the individual indexes were produced by typing entries onto single cards and filming these one line at a time with a special camera to produce a strip of film corresponding to a column of the index. Printing plates were produced from the film.[18]

The mid and late 1960s saw the development of an information processing system, partially funded by the National Science Foundation. This led to automation in the production of the indexes as well as the building of a database from which other newer search services could be derived. This effort was led at CAS by Fred A. Tate, who had been recruited as Assistant Director by Baker in 1961, by Russell J. Rowlett, Jr., who was appointed Editor in 1967, and by Ronald L. Wigington, who joined CAS as Director of Research and Development in 1968. They were instrumental in reorganizing the flow of materials so that abstracting and indexing were carried out simultaneously by document analysts. Additionally, mechanisms were set up so that data from the source documents were recorded only once in computer-readable form but then used multiple times depending on the type of output. Once the bibliographic, abstract, and indexing information had been extracted or derived from the source documents, it was organized, sorted, and formatted according to predetermined computer instructions.

Although this process for producing *CA* now seems fairly straightforward, the required technology did not exist in the last half of the 1960s. Consequently, it was necessary for CAS to develop a computer-directed system that did not require typographic instructions to be directly associated with the corresponding material in the database. Thus, in 1967 CAS developed a composition system incorporating the needs associated with an IBM Model 2280 film recorder. This technique was used to prepare the indexes until the early 1970s when other computer-controlled devices became commercially available. By the end of 1970, all indexes were computer composed, and the *Eighth Collective Index* (1967–1971) was prepared by merging the entries residing in the database from the ten semiannual volume subject indexes. Corrections and updates were included.

The computer has also been helpful in processing information that can be characterized precisely, thus aiding analysts in editing index entries. For example, headings in the *General Subject Index* are from a controlled list. That is, one term

EDIT TYPE	INPUT	OUTPUT	REASON
Spelling error	ffirst	ffirst???	Suspected spelling error
Abbreviation	preparation	prepn.	Approved abbreviation
Singular-plural	Knife	Knives	Plural form used for apparatus headings
Combined index heading	Sedatives	Hypnotics and Sedatives	Indexing policy
Cross-reference	Hydrophobia	Rabies	indexing policy

Figure 7. Examples of computer-assisted edits.

from several synonyms is selected as the heading—that most likely to be the one searchers have in mind—and cross-references are used from the others to that chosen one. Certain words are abbreviated consistently, and also, certain spelling errors may be detected. These kinds of operations, resulting in a diagnostic message to indicate a probable error, can be handled by machine.[19,20] Examples of assistance provided by the computer are shown in Figure 7.

Once the information is edited by an analyst, it is "released" into the database where it is selected by computer, using predetermined instructions, for the appropriate index, tape, or computer-readable file. After selection, the computer then converts it into the necessary form for publication, sorts and alphabetizes the index entries, justifies and hyphenates the text, formats and organizes the pages, and composes the material on film from which offset printing plates are made. Considerable savings in time, staff effort, and cost have been accomplished through the use of computerization.

Each of these improvements to the processing system has been implemented in a stepwise fashion. While it is desirable to shut down one complete system and replace it with one that is totally new, this is not practical. The production of *CA* cannot be interrupted for the length of time necessary to install, test, and debug a new system, nor could the staff be expected to absorb all the changes. A massive system change would also mean the retraining of large numbers of people as new assignments would be required. Stepwise implementation also permits efficient handling of problems that arise as a new collective period approaches. At any one time, data from two or three different volumes are being processed. Any changes in indexing policy between two collective periods must be smooth and as free from disruptions as possible since two sets of policies coexist for that period of time.

Presently, index entries, along with abstracts and keywords, are dictated by document analysts, transcribed clerically into computer-readable form, machine-edited, and displayed on cathode ray tube terminals for review by editors. Corrected entries again pass through maching edits and are displayed once more if errors are detected. Formerly, the entries were edited in hard-copy form, the changes were noted and rekeyed, and a new data sheet was ready to be reedited. However, rekeying at times produced new errors, which again had to be corrected. Online editing eliminated the several million sheets of paper that had to be handled, oftentimes by as many as four people. And it was necessary to keep track of each piece of paper to ensure that no access paths to a given document were lost.

These innovations have made it possible to have a document indexed, abstracted, and published in *CA* within 90 days of the date of publication of the document. This median currency decreases to 70 days when only journal articles are examined and to 60 days when the 750 "core" journals—those covered in *Chemical Titles*—are considered.

From the Index Guide

Acids
Studies of acids as a class are indexed at this heading. Specific acids are indexed at those specific headings.
See also *Titration, acid-base*
amphoteric—see *Amphoteric substances*
bile—*Bile acids*
bitter
see
α—*Bitter acids*
β—*Bitter acids*
boron—see *Boron acids*
carboxylic
see such headings as
Amino acids
Carboxylic acids
Fatty acids
Naphthenic acids
Resin acids and Rosin acids
Uronic acids
halogen—see *Halogen acids*
heteropoly—see *Heteropoly acids*
hydroxamic—see *Hydroxamic acids*
hydroximic—see *Hydroximic acids*
isopoly—see *Isopoly acids*
Lewis—see *Lewis acids*
nitrogen—see *Nitrogen acids*
ortho carboxylic—see *Ortho acids*
phenolic
see
Carboxylic acids
Phenols
phosphorus—see *Phosphorus acids*
sapogenin—see *Sapogenins*
sulfonic—see *Sulfonic acids*
sulfur—see *Sulfur acids*
tar—see *Tar acids*

From the Hierarchy

INORGANIC COMPOUNDS
Acids
• Arsenic acids
• • Thioarsonous acids
• Boron acids
• Halogen acids
• Heteropoly acids
• Hydrogen halides
• Hypohalous acids
• Isopoly acids
• Lewis acids
• Phosphorus acids
• • Peroxyphosphonic acids
• • Polyphosphoric acids
• • Superphosphoric acids
• Polyantimonic acids
• Sulfur acids
• • Thionic acids

ORGANIC COMPOUNDS
Acids
• Arsenic acids
• Boron acids
• • Boronic acids
• • • Thioboronic acids
• • Thioborinic acids
• Carboxylic acids
• • Amino acids
• • • Casamino acids
• • • Rozites
• • Benzenecarboxylic acids
• • α-Bitter acids
• • β-Bitter acids
• • Choleic acids
• • Fatty acids
• • Gibberellins
• • Hydroxamic acids
• • Hydroximic acids
• • Imidic acids
• • Naphthenic acids
• • • Thionaphthenic acids
• • Pyroligneous acids
• • Resin acids and Rosin acids

Figure 8. Hierarchical arrangement of index headings for acids.

Computerization also made it possible for CAS to begin producing other kinds of related subject indexes, not only for itself but also for the American Chemical Society (ACS). In 1961 CAS introduced *Chemical Titles* (*CT*), its first—as well as the world's first—journal to be totally composed and produced by computer. One of *CT*'s indexes is a keyword-in-context (KWIC) index of titles. In 1969, CAS, using the same computer programs used for the *CT* index, began to produce a similar KWIC index for the "Book of Abstracts" for the ACS national meetings. While the process is still completely automated, it has been upgraded over the years to reflect enhancements to data input, processing, and photocomposition.

In 1963, CAS introduced a *Keyword Subject Index* to the *CA* issues. Index entries are phrases, consisting of up to four or five terms, derived from the title and abstract for a document. There exists no relationship, however, between the keyword phrases and the more detailed index entries that appear in the corresponding *General Subject Indexes*. In 1972, CAS began to extract corresponding keyword phrases from its database for two of the ACS journals, *Inorganic Chemistry* and the *Journal of Organic Chemistry*, to prepare their annual indexes. Shortly thereafter, a third journal, the *Journal of the American Chemical Society*, also began to publish an annual index, again with entries extracted from the CAS database. Today, *CA Keyword Index* entries are extracted for 19 ACS primary journal indexes.

INDEXING IMPROVEMENTS

While maintaining this currency, users' recommendations for improvements are actively sought. Such requests and suggestions, derived from surveys, forums, letters, advisory boards, and interviews, representing industrial, academic, and governmental organizations, are the basis for research that has resulted in, for example, the formation of a hierarchical index to the *General Subject Index* headings.[21,22] Such hierarchical relationships have always been present, but headings have not been classified to reveal this structure.

Figure 8 shows an example of index headings that appear in the *Index Guide* and how these are related hierarchically to other headings. Note that broader term and narrower term relationships are not specifically identified in the main body of the *Index Guide* but appear only in the hierarchy. A given index heading may appear in more than one hierarchy. The main body of the *Index Guide* contains information not found in the hierarchies, and the hierarchies present relationships that, while present in the main body of the *Index Guide*, would have to be manually extracted and reformatted in order to bring them into focus. The example **Acids** is not complete; that is, other cross-references and hierarchy relationships exist but are not shown in the figure. Hierarchical levels are indicated by successively indenting to the right those terms that are more specific than those preceding them. Bullets (•) are used so users may identify those headings that are on the same hierarchical level, which are sorted alphabetically.

CAS has attempted to respond to users' needs by providing various kinds of assistance. One of these ways to aid users was to collect into one place all cross-references, indexing notes, details on CAS indexing policies, etc. This collection, the *Index Guide*, was introduced in 1968, to which a hierarchical list of the *General Subject Index* headings was added in 1976. CAS has also actively sought out suggestions on ways to improve the usefulness of its products and services. From 1963 through 1979, CAS Open Forums were held in conjunction with national meetings of the American Chemical Society. CAS has also obtained recommendations from members of the Editorial Advisory Board (1975–1982) and the present user councils. A workbook, "Using Chemical Abstracts Service Printed Information Services", was introduced in 1973 to lead users, step by step, through sample problems to show how search strategies may be developed and executed, and this was followed by the introduction of workshops conducted by CAS staff members. CAS has also presented and published papers[23-28] on selected subjects covered by *CA*.

CAS continues to upgrade and improve its indexing operations and publications while adhering to philosophies, developed over the years, that have assured quality and excellence. Having served as the Editor for 16 years, until 1982, Rowlett summarized the fundamental *CA* indexing philosophies in his 1983 Herman Skolnik Award address.[29] David W. Weisgerber, the present Editor of *CA*, continues to emphasize the building of a database of highest integrity—based on completeness, timeliness, and quality—from which both printed publications and computer-readable services are derived. While the online delivery of indexing information grows at a rapid pace, the need for printed indexes will continue worldwide.

REFERENCES AND NOTES

(1) Baker, D. B.; Horiszny, J. W.; Metanomski, W. V. "History of Abstracting at Chemical Abstracts Service". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 193–201.
(2) Crane, E. J., Ed. "CA Today: The Production of Chemical Abstracts"; American Chemical Society: Washington, DC, 1958; p 37.
(3) Bernier, C. L.; Crane, E. J. "Indexing Abstracts". *Ind. Eng. Chem.* **1948**, *40*, 725–730.
(4) Crane, E. J. "The Training of Chemists for Abstracting and Indexing". In "Training of Literature Chemists". *Adv. Chem. Ser.* **1956**, *17*, 16–21.
(5) Bernier, C. L.; Crane, E. J. "Correlative Indexes. VIII. Subject-Indexing vs. Word-Indexing". *J. Chem. Doc.* **1962**, *2*, 117–122.
(6) Crane, E. J.; Langham, C. C. "On-the-Job Training at Chemical Abstracts Service". *J. Chem. Doc.* **1962**, *2*, 199–204.
(7) "The First 75 Years of Chemical Abstracts Service". *CAS Rep.* **1982**, *June* (12), 3–9.
(8) Crane, E. J.; Patterson, A. M. "A Guide to the Literature of Chemistry"; Wiley: New York, 1927; p 79.
(9) Ibid., p 88.
(10) Ibid., p 89.
(11) "Oxford English Dictionary", 1971 ed.; Oxford University Press: New York, 1971.
(12) Tswett, M. *Ber. Dtsch. Bot. Ges.* **1906**, *24*, 316, 384.
(13) Crane, E. J. "Five Miles of Index". *Chem. Eng. News* **1947**, *25*, 1188–1189.
(14) Crane, E. J. "Why Indexers Turn Gray". *Ind. Eng. Chem., News Ed.* **1937**, *15*, 175.
(15) Troan, J. "An Adventure in Knowledge: The Story of the Chemical Abstracts Service"; American Chemical Society: Washington, DC, 1963; p 17.
(16) Mellon, M. G. "Searching the Chemical Literature"; American Chemical Society: Washington, DC, 1964; p 24.
(17) Tate, F. A. "Introduction", 14th Chemical Abstracts Service Open Forum, Chicago, IL, Sept 1970; American Chemical Society: Washington, DC, 1970.
(18) Wetsel, F. R. "Time Studies in Producing Subject Indexes for *Chemical Abstracts*". *J. Chem. Doc.* **1965**, *5*, 212–214.
(19) Nelson, R. D.; Hensel, W. E.; Baron, D. N.; Beach, A. J. "Computer Editing of General Subject Heading Data for *Chemical Abstracts* Volume Indexes". *J. Chem. Inf. Comput. Sci.* **1975**, *15*, 85–94.
(20) Pollock, J. J.; Zamora, A. "Collection and Characterization of Spelling Errors in Scientific and Scholarly Text". *J. Am. Soc. Inf. Sci.* **1983**, *34*, 51–58.
(21) "Subject Heading Hierarchies Published in Index Guide". *CAS Rep.* **1977**, *September* (6), 8.
(22) "Hierarchies of General Subject Headings". "*Chemical Abstracts* Index Guide", 1984 ed.; American Chemical Society: Washington, DC, 1984; Appendix I.
(23) Dickman, J. T.; O'Dette, R. E. "CAS as a Literature Source for Agricultural and Food Science and Technology". *J. Chem. Doc.* **1968**, *8*, 98–105.
(24) Dickman, J. T.; O'Dette, R. E. "Chemical Engineering Content of *Chemical Abstracts*". *Chem. Eng. Prog.* **1971**, *67*, 79–81.
(25) Langstaff, E. M.; Ostrum, G. K. "Access to Polymer Information in *Chemical Abstracts*". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 60–64.
(26) Moulton, C. W. "Fossil Fuels in *Chemical Abstracts*". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 83–86.
(27) Beach, A. J.; Dabek, H. F.; Hosansky, N. L. "Chemical Reactions Information Retrieval from Chemical Abstracts Service Publications and Services". *J. Chem. Inf. Comput. Sci.* **1979**, *19*, 149–155.
(28) Moody, R. L.; Zahm, B. C. "*Chemical Abstracts* as a Resource for Health and Safety-Related Chemical Information". *J. Chem. Inf. Comput. Sci.* **1980**, *20*, 12–14.
(29) Rowlett, R. J., Jr. "An Interpretation of Chemical Abstracts Service Indexing Policies". *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 152–154.