Symposium on User Reactions to CAS Data and Bibliographic Services. Concluding Remarks[†]

RUSSELL J. ROWLETT, JR.

Chemical Abstracts Service, The Ohio State University, Columbus, Ohio 43210

Rather than attempting to summarize in any way the great diversity of information and views presented here today, I would like to respond to some of the specific comments made and questions raised by the speakers and the audience. I would also like to expand somewhat on several topics that have come up in the discussion.

INDEXING NOMENCLATURE

One paper today dealt with reactions to *Chemical Abstracts* (CA) Index Names for the Ninth Collective Index (9CI) period. This nomenclature, which has been used in the CA Volume Indexes since Volume 76 (January–June 1972), is currently the topic of some controversy in the community.

It was absolutely imperative that Chemical Abstracts Service (CAS) adopt fully systematic naming practices for its semiannual and collective indexes. Had we continued to use the multitudes of trivial names that had developed over the years, it soon would be impossible for our users to search the indexes or for our indexers to name the rapidly growing number of new substances reported in the literature. Let me emphasize that CAS has not created any new nomenclature system. We have simply adopted the more fully systematic naming procedures already approved, accepted, and reported by the International Union of Pure and Applied Chemistry (IUPAC) and other nomenclature bodies. It is true that in a few specialty subject areas the IUPAC Commissions have not yet made up their minds about officially accepted naming practices. Since CAS must name new substances every day to meet production deadlines, we do have to create nomenclature in these subject areas, and we do so as best we know how. CAS is represented on most of the IUPAC Commissions, the ACS general and divisional nomenclature committees, and other such bodies. We know what the thinking of these groups is, and we try to select new names for which there is some precedent.

We do not expect CA Index Names to be used in normal communication by authors or lecturers or in industry—we don't really wish to see ethenylbenzene replace styrene on 55-gallon drums. However, in a highly organized alphabetical index in which we may employ only one name for each substance, a high degree of systematization in naming is absolutely essential. Prior to the 9CI period, hundreds of pages of instructions were necessary to instruct indexers how to name phenyl-substituted acyclic acids. This example was repeated manyfold for other types of substances. Someone commented on methylbenzene as the current index name for toluene. If you add a methoxy group to methylbenzene, what is the name of the resulting structure? Is it methylanisole or methoxytoluene? In today's CA Volume Indexes it is methoxymethylbenzene. How many of you know the difference between pyrocatechuic acid and protocatechuic acid? How about α-, β-, and γ-resorcylic acids? They are all dihydroxybenzoic acids. In the current

CA Chemical Substance Index you will find them all listed together as dihydroxybenzoic acids. We feel that this arrangement that brings together related compounds within the index has merit.

I really do not believe that chemists get through school without understanding the basic principles of systematic nomenclature. To use the 9CI names, you no longer need to remember trivial names with the exception of some ring names, the names of the stereoparents, and a few others. For those who do not like 2-propanone in place of acetone and question why we did not go absolutely all the way in adopting systematic names for every substance, it was because the Editor just did not have the intestinal fortitude to go all the way. The only trivial names retained, other than those that include sterochemical information, are acetic, formic, carbonic, and benzoic acids, phenol, and hydrazine. The IUPAC has no way of handling hydrazine unless replacement nomenclature is employed. The others are so common and have so many derivatives that it would have been very difficult to convert them all to fully systematic names in the time limits within which we had to work. We also surveyed everything that has been suggested for systematizing ring names, but have not come up with a good way to do it. The trivial names for stereoparents include the third structural dimension in their names, and those who work in sterochemistry have not devised good ways to systematize all of these names. When there are 17 asymmetric centers in a molecule, it makes for a rather long and complex systematic name. So we have retained cholestane, pregnane, and similar names, and have no plans for changing them until the science gives us some better ideas.

I am certain if you will work with the 9CI names for a while, you will come back and tell me that they are easier to search and to generate. This has been our experience at CAS and the experience of others who work daily with large numbers of substances. Once you see the logic and understand the value of systematic names in a highly ordered index, I think you will defend the 9CI nomenclature.

USER EDUCATION

Another of the papers concerned a workshop seminar on the use of CAS services conducted as part of a National Federation of Abstracting and Indexing Services program in 1973 and the accompanying experimental workbook. The statement was made, I believe, that there had been competition in preparing workbooks for the use of printed and computer-readable products, and since the workbook for printed services was completed first, the printed services were featured on the NFAIS program. This is an entertaining story, but not quite true. What our users need most urgently is guidance and instruction on how to address the technical content of the CAS data base. This content is the same whether it is in printed or computer-readable form, and users must generate similar profiles whether they address it in printed or computer-readable form. Since more people are familiar with this content in printed form than in computer-readable form, the workbook for printed services was introduced first.

This leads me to the general topic of user education at CAS. CAS has recognized always the need to instruct users in the use of its publications and services. We have not in the past, however, had the qualified staff, the time, or the resources to do as much in this area as we wished or as much as we knew was needed. Only recently have we been able to begin to build the resources to do something directly about user education.

Most of our staff members are experts in technical subjects, in languages, and in how to put indexes together. Few of them really are familiar with how these indexes are used in the environment in which our users work. The introductions to the CA indexes, for example, used to explain how the indexes were put together, not how to search them. Since the late 1960's we have attempted to reorient these introductions so that they tell the user how to find the information he seeks. This has required considerable reorientation in the thinking of our staff. We have not yet completely accomplished all we need to do in this area.

Recently CAS established a focal point for user education activities in its Marketing Department. The user education program that we currently are developing has two main elements: the workshops and problem workbooks that were referred to earlier, and a package of user aids that have been described at CAS Open Forums and referred to several times here.

The user aids represent an attempt to derive from the computer data base frequency lists, comparisons, statistics, and the like that have not been available before and from which users can get a better concept of the technical content of the data base and how to address it. While these aids are usually described in the context of computer-readable products, they are equally useful in searching the printed indexes.

I should also note that much of the impetus for the CAS workbooks came from the audiovisual course, "The Use of *Chemical Abstracts*," prepared by Dr. O. Bertram Ramsey of Eastern Michigan University and distributed by the American Chemical Society's Department of Educational Activities. We feel that Dr. Ramsey's course is an excellent general introduction to the use of CA. However, it required two years after the final editing to produce the audiovisual package. Changes do occur in a two-year period, so the present audiovisual course does not in all respects reflect CA and other CAS publications and services as they are today. There are also several other audiovisual guides to using CA that were produced outside of the United States, and a very excellent book on the subject which, unfortunately, has been published only in Japanese.

User education is a costly undertaking. Workbooks are not prepared cheaply, and it requires top senior staff members to develop the workbooks and workshops. It has required time to develop the CAS processing system to the point where we can afford to assign responsibilities in the area of user education to senior staff. We feel we have now reached that point.

Let me emphasize that the CAS user education program is experimental at this point. The workbook is not yet in a condition in which we can make it generally available. We are gradually improving it as we gain experience in it's use. We hope that we may someday be able to convert it into a stand-alone text that might be useful in a college chemical literature course (there are a number of people interested in seeing us do that), but we do not know whether it is possible. We are finding in conducting workshops for various groups that often each group requires a different approach. We feel presently that we need to take a workbook that contains a relatively broad selection of problems into an organization and discuss those that relate most closely to the subject areas with which the audience is most familiar.

In any event, we do now have the ability to conduct workshops for a limited number of organizations each year.

If you are interested in our providing such a workshop for your organization, please contact the User Education Group in the CAS Marketing Department. The procedures and costs can be worked out.

USER FEEDBACK

One problem CAS faces in attempting to instruct its users is that of finding the users. They are not on our lists of subscribers. Most of our services are mailed to librarians, purchasing agents, or book dealers.

This makes it difficult also to obtain meaningful feedback from users. The Editor receives a few letters from users, but they represent only a minute percentage of the total body of users. We once had a Market Research Department that sent out many questionnaires and even got a few replies, but the replies were from those on the subscription lists, not from those who really use CA.

In recent years we have taken advantage of opportunities for talking to users through open forums at ACS meetings, through a number of ACS committees, and through seminars around the country, meetings of associations of information centers and users of computer-readable files, and symposia such as this. We hope to be able to do more of this.

Very recently we created a CAS Editorial Advisory Board in the tradition of such boards for other American Chemical Society journals. This new board functions separately from and in addition to CAS's other, more broadly constituted Advisory Board, which includes representation from the library, information science, data processing and business communities. Establishment of the Editorial Advisory Board is another step in our efforts to maintain the highest possible standards of quality in the technical content of the CAS data base. Members of this new board represent primarily users, or large groups of users, of CAS publications and services. I urge you to make your views on matters of technical content known to the members of this board.

CA CONDENSATES AND KEYWORD INDEXES

The statement was made in one paper today that the CA weekly indexes come from *CA Condensates*. I would like to correct this misunderstanding. There is only one, unified CAS data base, and the technical contents of all CAS services—both printed and computer-readable—are derived from this data base. *CA Condensates* is a copy of certain data elements (abstract headings, keywords, and CA section numbers) selected from that data base. The weekly CA Issue Keyword Indexes represent a printed subset from the same data base. They are not produced from *CA Condensates*.

Also with regard to *CA Condensates*, we are asked frequently whether or when we plan to add various kinds of data to this computer-readable file. One particularly frequent question concerns Registry Numbers. At the moment, we have no plans to add data elements to *CA Condensates*. My personal view is that if we add new types of data to *CA Condensates*, it will no longer be *CA Condensates*, but a different service. We are, however, always interested in hearing users' ideas about our products and always willing to consider them.

While we are on the topic of keyword indexes, I would like to try to clarify a few things about the nature of these indexes and some of the ways in which they differ from the CA Volume Indexes. Issue Keyword Indexes were added to CA in 1963. The issues had grown so large at that point that we felt it was necessary to provide some additional means, beyond the organization of abstracts in subject section, for users to locate abstracts on particular subjects. These keyword indexes started out as "quick and dirty" in-

dexes, and they essentially still are, though we have made some improvements in the use of abbreviations and eliminated some repetition and some of the duplication inherent in permuted keyword phrases. The keyword indexes definitely were not designed for computer searching, and we seek your experience in using keywords in computer searching to aid us in improving their usefulness for this purpose.

I would like to emphasize also that there is no connection between the content of the weekly Issue Keyword Indexes and that of the six-month Volume Indexes. For one thing, we do not include specific chemical substances in the weekly keyword indexes, though we do indicate in a generic way the classes of substances present. For example, the keyword "phenol" may indicate the presence of information on phenol itself or on substances of the general class of phenols. There is no attempt even to index the same concepts in the Issue and Volume Indexes. Under our recently instituted unified document analysis process, keywords and Volume Index entries are generated by the same staff member, and I am sure they tend to select the same basic indexable ideas for both indexes. We do not direct them to do so, however, and we cannot afford to produce an Issue Index as complete or as detailed as the six-month Volume Index.

The keyword indexes reflect, for the most part, authors' terminology. Users support the basic philosophy that abstracts and Issue Indexes should continue to reflect the terminology employed by the author. Searches of such terminology within the keyword indexes in combination with searches of the Volume Indexes, which use, for the most part, a controlled indexing vocabulary, provide a dimension of retrieval beyond that possible with either index used alone.

OTHER QUESTIONS

One question raised earlier concerned the procurement of copies of Soviet documents that are abstracted in CA. CAS in the past supplied copies of these documents from Columbus through a coupon service. We discontinued this service after the Soviet Union became a party to the Universal Copyright Convention in May 1973. The Soviets in turn have stopped copying CA; they previously had been reproducing 800 to 900 copies for distribution in the Soviet Union and Eastern Europe. We now have more than 100 new paid subscriptions from the COMECON countries, and this is a very genuine help in the support of CA.

The Soviet recognition of copyright also precludes our publishing in CA translations of copyrighted abstracts from *Referativnyi Zhurnal*. We are, however, publishing the titles of these abstracts in CA (we published about 5000 of them near the end of 1974) in the belief that the scientific community in the Western nations would prefer to have a record of their existence.

A question also was raised about the availability of the contents of Eighth Collective Index in computer-readable form. All of the contents of the CA Volume Subject Indexes from 1967 forward are available in computer-readable form. The complete Volume Subject Indexes produced during the Eighth Collective period were issued in computer-readable form as the CA Integrated Subject File. These are extremely large files. Current index data are available in files of more manageable size through the CA Subject Index Alert, which is issued biweekly. Each issue of this file contains all of the index entries that have been edited, verified, and prepared for publication in the Volume Subject Indexes during the preceding two weeks.

Another question concerned the possible eventual availability in computer-readable form of CA abstracts dating back to 1907. All current CA abstracts will be computer composed by mid-1975, and all abstract text thus will be available in computer-readable form from that time forward. We have no plans, however, to enter earlier abstracts in the computer data base. Rekeying that amount of material would be a horrendous and extremely costly undertaking. We do hope someday to be able to create computer records for chemical substances indexed in CA prior to the beginning of the CAS Chemical Registry System in 1965, but with current funding restrictions, this is not likely to happen soon.

Finally, the results of one of the surveys reported today, I believe, questioned the timeliness of the appearance of the CA Volume Indexes. The question may have really concerned the computer-readable form of these indexes, but let me comment on the timeliness of the printed Volume Indexes anyway. Not too long ago these indexes were appearing about 13 months after completion of the volume of abstracts they covered. Last year they followed the volume of abstracts by just over six months. This year we are computer selecting subject material in April for the index covering the last six months of 1974. Our goal is to make the Volume Indexes available within three months of the end of the period they cover. We think this is attainable with our processing system. Three months is the approximate time needed to print, collate, bind, and mail the indexes.

DISCUSSION

Q. What happened to the Hetero-Atom-in-Context Index? It was a useful index even though we might not have used it often.

A. The Hetero-Atom-in-Context Index was produced for several CA volumes, and we received no comments about it whatsoever. In view of the lack of comments and because it was very costly to produce, we discontinued it. I believe you are only the third person to inquire about it. I too believe that it was a useful index, but I have no facts to support its value to users. We still have the computer programs that were used to create it and could recreate it if there is a genuine demand. At present we do not see that demand.

Q. Wouldn't it be better to poll your users before eliminating a service such as the Hetero-Atom-in-Context Index or making major changes?

A. It might be if we could address our users. I've already commented on the difficulty of doing so. We never make a major change in a product without first informing the American Chemical Society Board of Directors and its committee on Chemical Abstracts Service, the ACS Council Committee on Publications, and the CAS Advisory Board. We also announce the change in *Chemical & Engineering News* and describe it in the introduction to the appropriate index or the introduction to the first issue of the CA volume. We don't know what else we can do.

Q. I understand the desire to group similar compounds at one point in the Chemical Substance Index, but it seems to me that you got just the opposite result with one of the nomenclature changes for the Ninth Collective period. I refer specifically to the indexing of ammonium compounds, which used to be in one place in the indexes. Now it is necessary to search a series of amine homologs and look many places in the indexes because of a change in the order of precedence of chemical functions.

A. I personally argued strongly for continuation of the ammonium compound heading in the Chemical Substance Index, but I was out-argued by experts both at CAS and in the IUPAC and elsewhere. They argued that these substances are derivatives of the amines and, therefore, belong with the amines. They held that we were scattering compounds under our previous practice because the most important part of the molecule was the amine.

Q. Is there going to be a new edition of the "Ring Index"?

A. We intend to replace the "Ring Index" with a publi-

cation known as the "Parent Compound Handbook". The editorial work for this publication is done, but the programming necessary to computer compose the necessary structure diagrams has not been completed. It was deferred in favor of some more pressing projects. Very few people have inquired about a new edition of the "Ring Index". The publication of drawings with locant numbering for new rings in each Volume Index appears to be sufficient for most users at the moment.

Q. You mentioned a change in abstracting policy regard-

ing Russian publications. Have you also changed indexing policy?

A. My statement regarding abstracts of Russian publications applied only to those abstracts that we formerly translated from the *Referativnyi Zhurnal*, not to all Russian publications. For those references, we now publish only the title, and we index only from the title. This represents only a small fraction of our Russian coverage, however, and there have been no basic changes in indexing policies.

Terse Literatures. II. Ultraterse Literatures

CHARLES L. BERNIER

State University of New York at Buffalo, Buffalo, New York 14260

The creation of Ultraterse Literatures (ULs), demonstrated in this paper by reading the set of 30 Terse Conclusions (TCs), can be a part of generalization and of the creation of new knowledge and new research projects. Computers can probably be programmed to collect indexed TCs under index headings and under categories of index headings, and then to print out only those TCs under headings and categories that exceed a figure of merit or a specified growth rate. From the anticipated printout of important headings and categories of TCs, those people knowledgeable in the subject field can probably create ULs. TC: ULs with condensations to 1/1,000 or fewer the number of words in original works can probably be written from sets of TCs selected (possibly by computer) on the basis of the number of TCs, rate of growth of the set, or other figure of merit.

INTRODUCTION

Methods for keeping current in science and for putting to use what is read have become inadequate in the past half century.[1] Three of the principal methods for dealing with these inadequacies have been (1) specialization, (2) condensed surrogation, and (3) skipping. (1) Specialization has become overspecialization; closely related material that cannot be read for lack of time may be found later to be important. (2) Condensed surrogates (writings, condensed from the original literature, that can be used in place of the original works; examples are abstracts, annotations, and extracts) are now discovered to be delayed; too wordy to read, assimilate, integrate, remember, and use; and to be unorganized. Nearly all surrogates are scattered in space and time, and by language, and must be assembled and translated for use. (3) Skipping has never been satisfactory because the part skipped may contain essential information. Whether the skipping is done by ignoring articles, authors, books, languages, patents, or index entries in manual search, or is done by the use of the Boolean negative or too many Boolean products simultaneously in a correlative search[2] is immaterial; essential information can be lost.

Support for the existence of the problem of too much to read is to be found in hundreds of statements made by responsible individuals during the last 50 years. If it is possible (and if there is ready access to articles, language ability, and time) to read[3] two research papers (about 4,000 words) per hour, and reading is continued 365 days a year, then:

• The 10,000 articles in either oncology or cardiovascular research (as examples) estimated to be published each year would take 14 years to read.

• The 2,000,000 biomedical articles[4] per year would take 27.4 centuries to read; and

• The 413,000 papers and patents[5] of chemistry and chemical engineering estimated to be published in 1975 would take 5.66 centuries to read.

A third source of support for the existence of the problem comes from lags in use of research results. One illustration is:

The energy crisis, reported to be approaching for at least 4 decades, caught those in a position to contain the crisis unprepared to take effective action. This occurred despite a literature[6] that carries many tested ways to generate and conserve energy.[7]

The technical literature, well tested and reliable, remains unused, too-little used, or used only after years of delay. Some lags are deadly, some prolong suffering, and some are costly, e.g., $22.4 billion lost annually from heart disease through income not earned, diagnosis, therapy, hospitalization, and burial (in the U.S. in 1962).[8] For 1975, the cost is estimated to increase to $36.6 billion if we assume an increase from 600,000 to 700,000 deaths per year plus total inflation of 40%.

REPACKAGING

Abstracts, condensing the primary literature to about one-tenth the wordage, have been a fairly effective way of helping solve the problem of too much to read. However, there are still too many words in abstracts that are irrelevant to the immediate purpose of the searcher. Extracts that condense to about one-fifth the number of words present more irrelevant words than do abstracts. Handbooks of