Searching *Chemical Abstracts* vs. *CA Condensates*†

ROBERT E. BUNTROCK

Standard Oil Company (Indiana), Amoco Research Center, Naperville, Illinois 60540

Received April 3, 1975

As of March 1975, as much as four and one-half years of *Chemical Abstracts* indexes and on-line accessible *Chemical Abstracts Condensates* can be compared. Although combined searches of both data bases are the most effective, examples are shown in which it is more practical and efficient to search *CA Condensates*. CHEMCON and CHEM7071, the on-line versions of *CA Condensates* loaded at System Development Corp. (SDC), are compared with CA Indexes.

Although most *Chemical Abstracts* users are familiar with the *Chemical Abstracts* Volume or Cumulative Indexes, most are probably not very familiar with *Chemical Abstracts Condensates*. *CA Condensates* is a machine readable bibliographic and indexing aid to *Chemical Abstracts*[1] and appears in its most familiar form as the index in the rear of the weekly issue of *Chemical Abstracts*. Although *Chemical Abstracts* is the data base that is being used, both the Indexes and Condensates are defined as data bases in this presentation.

When Condensates became commercially available from Chemical Abstracts Service, searching services and information centers began to offer Selective Dissemination of Information (SDI) services from the weekly tapes by sequential batch-mode processing of user profiles. Some centers accumulated back files of the tapes and offered retrospective searches, again processed batch mode. Except to those who had direct access to the tapes, Condensates still remained mysterious. The resrospective searches were rather expensive and "browsing" was impractical. As for the weekly indexes, searches involving coordination of two or more concepts or substances rapidly became difficult and impractical. It was often easier to look up all of the citations under a given term, or terms, and determine relevancy from the abstracts. Other than the entry point by which it was found, the indexing of any abstract was unknown.

The availability of Condensates for interactive on-line searching changed a lot of things. "Browsing" is now possible and the complete indexing can be found for any item. The annoying time lag between search entry and receipt of results—hours to days for batch systems—can now usually be reduced to minutes. Unlike batch systems, incorrect or inadequate search strategy can be determined immediately. The apparent but often debated advantages of random-access on-line searching compared to sequential batch searching methods[2] seem to apply.

Increased availability has also raised additional questions of the true worth of *CA Condensates*. The indexing of *CA Condensates* is obviously less systematic and less exhaustive than that of the CA Volume or Cumulative Indexes. As more of the two data bases overlap, one must often decide if duplicate searches are worthwhile. Some comparisons of *CA Condensates* with other data bases have been reported, and *CA Condensates* generally appeared to be the inferior data base. Buckley cited an example where the CA Index provided more unique references (for the manufacture of Terramycin) than did *CA Condensates*.[3] The University of Georgia Search Center found the performance of *CA Condensates* to be inferior to the *CA Integrated Subject File*[4] (ISF, a machine readable form of the CA 9th Collective Index). Different forms of *CA Condensates* have also been compared. A comparison of *CA Condensates* as searched on-line with manual searching of the weekly issue indexes is reported by Michaels,[5] and a comparison of the two commercial on-line loadings of *CA Condensates* is reported by Prewitt.[6]

At the Amoco Research Center, we too have compared search results for *CA Condensates* and the CA Indexes, and have found examples for the superiority of *CA Condensates*, superiority of the Indexes, and virtual equivalence. As of March 1975, at least four years of *CA Condensates* and the published Indexes could be compared (Vols **72–79,** 1970–1973). For author and general subject searches, Vol 80 (Jan–June, 1974) could also be compared. *CA Condensates* is usually disadvantageous for searching specific compounds and Buckley's example[3] is typical. However, we found that other types of search examples (corporate author, patent assignee, personal author, generic/specific compound, and reaction class) illustrate the comparative strengths of *CA Condensates* for practical searching. In the following examples, CHEMCON and CHEM7071 (System Development Corp.) are the on-line loadings of *CA Condensates* that were used.

SEARCH EXAMPLES

Corporate Author. As a corporate author example, searches were run to determine how Chemical Abstracts Service indexes itself and its products. The on-line strategy is as follows:

((CHEM OR CHEMICAL) AND (ABSTR OR ABSTRACT# OR CONDENSATES)) OR CAS OR ((CHEM/CS OR CHEMICAL/CS) AND (ABSTR/CS OR ABSTRACTS/CS))

The hash mark or pound sign now represents only single character substitution in ORBIT III, SDC's new operating program, and is usually explicit for a term and its plural. As there are only 16 citations under "Chemical Abstracts Service" in the Vol. 72–80 Indexes and many more are retrieved by the above on-line strategy, CAS and its products must appear as subheadings under other main headings. Under "Information Science" in the CA Index Guide, "see also" items include "Indexing," "Literature," and "Periodicals." Scanning all four headings for CAS-oriented items produced enough additional references to indicate the data bases are equally effective (Table I) (for arbitrary reasons, citations referring only to *Chemical Titles* or *Chemical-Biological Activities* were excluded from both outputs).

† Presented in the Symposium on "User Reactions to CAS Data and Bibliographic Services," 169th National Meeting of the American Chemical Society, Philadelphia, Pa., April 7, 1975, Abstract CHLT 7. Also presented at the 66th Annual Conference of the Special Libraries Association, Chicago, Ill., June 1975.

Table I. Comparison of Citation Retrieval for "Chemical Abstracts"

	Unique citations	Citations in common
CA Condensates	16	28
CA Indexes	17	

Note that the number of unique references retrieved from each data base made the dual search worthwhile.

Several reasons can be given for these performances. In the Indexes, additional references concerning CAS or *CA Condensates* could probably be found as subheadings under additional main headings, but the statistics above reflect probably the best practical search. *CA Condensates* often indexes too generally, as can be shown by "recycling" the citations unique to the Index into *CA Condensates*. Frequently these citations were indexed to "Information Service." Retrieval for the on-line search was increased by use of the "Corporate Source" portion of the search. In the Indexes, on the other hand, only patents are cited for an institutional author. In CHEMCON, the "Corporate Source" field consists of single words, derived from the Corporate Author, Patent Assignee, and Work Address fields, with a "CS" role attached. It is therefore valuable in determining the institutional origin of any kind of document and can, of course, be coordinated with other search elements including personal authors and keywords.

"Corporate Source" (currently available for CHEMCON only) aids search efficiency even if confined to a patent assignee function. In the Index, the patents for each organization are printed in order of ascending abstract number and not by subject. For organizations that issue large numbers of patents, the resulting task of scanning titles (which are often inadequate) can be formidable. In searching on-line, keywords and title words can be coordinated with the organizational author. In a test of the recall of Condensates, all 155 patent citations for "Standard Oil Co. (Indiana)" found in the Indexes were also retrieved on-line for Vols 76–80. On-line searching is even more valuable in situations where, unlike Indiana Standard, patents are issued to any of several divisions or subsidiaries of an organization with different names, e.g., several large Japanese companies.

Personal Author. Personal author searches can often be performed faster on-line because, unlike in the Indexes, the author's name can be searched directly regardless of the order of appearance of names for multiauthored articles. A good example is a search for articles by Professor Paul Schleyer of Princeton University. Use of the "Neighbor" command for P. Schleyer produces the following listings from Condensates (1972–Feb 1975):

Schleyer P. R.	Schleyer P. Von Rague
Schleyer P. V. R.	Schleyer Paul R.
Schleyer P. Von R.	Schleyer Paul V. R.
	Schleyer Paul Von R.

all of which can be retrieved by "SCHLEYER P:" (truncated). All of the references can be checked on-line in a few minutes by printing further bibliographic details, and all are authored by the author of interest. Although Paul Schleyer is listed in the corresponding Volume Indexes only two ways, he is not the first named author for many papers and many cross references must be looked up under the appropriate initial author. Even when all abstract numbers have been found in the Indexes, one must look up the abstracts even to determine further bibliographic information.

Compounds. For searching of specific compounds, the *Chemical Substance Index* appears to be superior to *CA Condensates*, mainly because of the greater depth of indexing in the Indexes and the more generic approach taken for those items that are indexed in *CA Condensates*. Although the ISF/*CA Condensates* comparison study showed no significant differences between the *Chemical Substance File* and *CA Condensates* for exhaustive indexing,[4] many instances can be found where *CA Condensates* indexes no compounds (or classes of compounds) at all in the area of interest, or indexes noticeably fewer than the CA Indexes. The generic nature of *CA Condensates* produces precision problems for specific compounds, both with the usual loss of locants (e.g., all xylene isomers under "xylene") and with indexing to the compound class (e.g., xylenes under "alkylbenzenes" or "alkyl benzenes"). However, this generic indexing can be used to advantage if the search request is somewhat generic as well as specific. For example, a user interested in a reaction of *p*-xylene may be interested in that reaction of the other xylenes as well, or even of alkylbenzenes in general. In addition to examples shown above, ring stems and other name stems usually appear alone. For example, BIPHENYL# (biphenyl or biphenyls) will retrieve most citations from *CA Condensates* that deal with the biphenyl ring system.

Table II. Comparison of Citation Retrieval for "Nitrones"

	Unique citations	Citations in common
CHEM7071	58	34
8th Collective (1970–71)	19	
CHEMCON	113	22
Volume Indexes (1972–June 1974)	5	

For those name stems that appear as primary headings in the Indexes, (e.g., biphenyl) searches for compounds as entities and within their class can be done, although tediously, by scanning all the headings. However, some compound class names or name stems do not often appear (or never appear) as main headings for specific compounds, e.g., "ether", "ketone". Another example is "nitrone" which formerly appeared as a main heading but is now indexed (specific examples, 9th Collective Index) under the name of the amine followed by "N-oxide" in uninverted fashion. Only general studies appear under the heading "Nitrones" (in the General Subject Index). Citation statistics for "nitrone or nitrones" are shown in Table II, divided into the 8th and 9th Collective indexing periods. The indexes are obviously less practical for a general search.

Often, generic searches of inorganic compounds are requested. For example, if various aspects of sodium hydroxide, sodium bicarbonate, or sodium carbonate are to be searched, simple coordination of the aspects with "sodium" may be sufficient for an on-line search. In the Indexes, searches must be made for the three compounds as main headings and as subheadings inverted under the aspects in question. If any sodium compound is of interest, the on-line search is even more efficient. For general searches of anions (e.g., "sulfate" or "sulfates"), searching *CA Condensates* is again more efficient than searching the Indexes, as the Index headings are usually for very general studies only.

Information requests often require a general yet specific searching capability. Organometallics are very often advantageously searched in *CA Condensates*. For example, a general search for aspects of lead alkyls can be partially constructed for on-line searching by means of the following "or" clause:

ORGANOLEAD# OR TETRAALKYLLEAD# OR TETRAARYLLEAD# OR TRIALKYLLEAD# OR TRIARYLLEAD# OR DIALKYLLEAD# OR DIARYLLEAD# OR ALKYLLEAD# OR ARYLLEAD#

Table III. Comparison of Citation Retrieval for "Decarbonylation"

	Unique citations	Citations in common
CHEM7071	18	44
8th Collective (1970–71)	21	
CHEMCON	15	61
Volume Indexes (1972–June 1974)	17	

In addition, tetramethyllead, trimethyllead, tetraethyllead, etc., terms may be added if appropriate. Finally, recall can be increased even further, with a loss in precision, by constructing an "or" clause of the prefixes of the terms above which is then coordinated with "lead". In the Indexes, such compounds are inverted under headings such as plumbane, for example. Such listings can go on for many pages in a Cumulative Index. For such searches, the use of *CA Condensates* is more practical.

Reaction Class. Finally, searching of processes or reaction classes produces mixed results for effectiveness for the two data bases. However, the recall of the combined search is often appreciably higher than that for either data base alone. In the case of "decarbonylation" more unique citations were retrieved from the Indexes than from *CA Condensates,* but each file appreciably enriched the combined search. In this case, "recycling" of some of the unique Index citations into *CA Condensates* prompted an additional on-line search on "photodecarbonylation" which improved the performance of *CA Condensates* by five citations. The final results are shown in Table III.

CONCLUSIONS AND RECOMMENDATIONS

With the possible exception of searching for specific compounds, the use of *CA Condensates* enhances the quality of *Chemical Abstracts* searches and, in fact, may at times be the best practical searching method. Reasons for deficiencies in *CA Condensates* can be determined by searching for specific items found elsewhere (by abstract number, authors, etc.) and printing the entire indexing record. Such "recycling", as mentioned in the CAS/Corporate Author and reaction class examples, is a valuable searching exercise for many files. The examples were not meant to be exhaustive, but they are typical and can be used to predict the outcome of other searches. In general, duplicate searches should be performed both in the Indexes and in *Condensates* even within the period of overlap. In spite of some deficiencies, *CA Condensates* has a high enough demonstrable value so as to prompt the following recommendations:

(1) Retain the back files of *CA Condensates,* either searchable on-line *in toto* or available for on-line "scanning" and "off-searching". This is the only record of indexing for users who do not have direct access to the tapes.

(2) Promote the study and eventual on-line availability of the ISF or suitable ISF/*CA Condensates* merged files. Hopefully, it will be possible to use the inherent substructure searching capability of Ninth Collective Index names.

(3) As an interim measure of lasting value, promote the on-line availability of the new tape services of CAS[7] that are analogous to CBAC and POST. All of these files contain CAS Registry Numbers and allow for free-text searching of the abstract.

(4) Promote the on-line availability of the *CA Patent Concordance* or, better yet, the on-line availability of a combined concordance or consortium of concordances from CAS, API, IFI, and Derwent.

LITERATURE CITED

(1) "Information Tools 1975. Book One," Chemical Abstracts Service, Columbus, Ohio, 1975, p 18.

(2) Lancaster, F. W., and Fayen, E. G., "Information Retrieval On-Line," Melville Publishing Co., Los Angeles, Calif., 1973.

(3) Buckley, J. S., Jr., "Planning for Effective Use of On-Line Systems," 168th National Meeting of the American Chemical Society, Atlantic City, N.J., Sept 1974, Abstract CHLT 14.

(4) Zipperer, W. C., Park, M. K., and Carmon, J. L., "The CA Integrated Subject File. II. Evaluation of Alternative Data Base Organizations," *J. Chem. Doc.*, **14,** 15–23 (1974).

(5) Michaels, C. J., "Searching CA Condensates On-Line Vs. the CA KWIC Indexes," 169th National Meeting of the American Chemical Society, Philadelphia, Pa., April, 1975, Abstract CHLT 6; *J. Chem. Inf. Comput. Sci.* **15,** 172 (1975).

(6) Prewitt, B. G., "Searching the *Chemical Abstracts Condensates* Data Base via Two On-Line Systems," 169th National Meeting of the American Chemical Society, Philadelphia, Pa., April 1975, Abstr. CHLT 8; *J. Chem. Inf. Comput. Sci.*, **15,** 177 (1975).

(7) "Four New Computer-Readable Products. 1. Energy. 2. Materials. 3. Food and Agricultural Chemistry. 4. Ecology and Environment," Chemical Abstracts Service, Columbus, Ohio, 1975.