LITERATURE CITED

(1) Farrell, C. D., Chauvenet, A. R., and Koniver, D. A., "Computer Generation of Wiswesser Line Notation," *J. Chem. Doc.* **11**, 52 (1971).

(2) Feldmann, R. J., and Koniver, D. A., "Interactive Searching of Chemical Files and Structural Diagram Generation from Wiswesser Line Notation," *Ibid.,* **11**, 154 (1971).

(3) Granito, C. E., Roberts, S., and Gibson, G. W., "The Conversion of Wiswesser Line Notations to Ring Codes. I. The Conversion of Ring Systems," *Ibid.,* **12**, 190 (1972).

(4) Hyde, E., Matthews, F. W., Thomson, L. D., and Wiswesser, W. J., "Conversion of Wiswesser Notation to a Connectivity Matrix for Organic Compounds," *Ibid.,* **7**, 200 (1967).

(5) Hyde, E., and Thompson, L., "Structure Display," *Ibid.,* **8**, 138 (1968).

(6) Lynch, M. F., "Conversion of Connection Table Descriptions of Chemical Compounds into a Form of Wiswesser Notation," *Ibid.,* **8**, 130 (1968).

(7) Miller, G. A., "Encoding and Decoding WLN," *Ibid.,* **12**, 60 (1972).

(8) Petrarca, A. E., Lynch, M. F., Rush, J. E., "A Method for Generating Unique Computer Structural Representations of Stereoisomers," *Ibid.,* **7**, 154 (1967).

(9) Thomson, L. H., Hyde, E., and Matthews, F. W., "Organic Search and Display Using a Connectivity Matrix Derived from Wiswesser Notation," *Ibid.,* **7**, 204 (1967).

(10) Vander Stouw, G. G., Naznitsky, I., and Rush, J. E., "Procedures for Converting Systematic Names of Organic Compounds into Atom-Bond Connection Tables," *Ibid.,* **7**, 165 (1967).

A Microfilm Index to *Chemical Abstracts*

F. ROBINSON

CIU—Development, Imperial Chemical Industries Ltd., Imperial Chemical House, Millbank, London SW1P 4QG, England

To improve access to the recent *Chemical Abstracts,* a cumulative quarterly index, based on the keyword phrases, has been produced in microfilm form. By the use of the *CA Condensates* tapes and Computer Output Microfilm (COM) the index is available soon after the end of each quarter. Abstract titles are included in the index, thus increasing its value as a working tool.

Chemical Abstracts plays a fundamental part in the information services of most chemical companies. The printed volumes are scanned by most information units for current awareness and are also searched retrospectively on specific topics. The continuing growth in the number of items covered[1]—although it emphasizes the importance of CA as an information source—makes its use more difficult. The possibility of handling a reasonable proportion of the relevant information in original form is increasingly beyond the resources of any single company.

In common with many companies, ICI has considered the use of the various tape services offered by Chemical Abstracts Service and, in particular, *CA Condensates.* The first study was of the use of Condensates for retrospective search, and the problems of developing a computer system for in-house use were considered. These are complex, mainly because of the size of the data base, and it was felt that experience with Condensates for SDI using existing programs was necessary before implementing a system capable of retrospective search. The programs written by the National Research Council of Canada were chosen for this service; the NRC SDI system has been described elsewhere.[2]

However, the *CA Condensates* tapes could be processed relatively simply to provide, every few months, a cumulative index to ease the problems of retrospective search. The development of this tool and its use are described here.

The problems of searching *Chemical Abstracts* itself are well-known. The CA subject indexes are admirable for those volumes for which they have been issued, but real difficulties arise with the more recent volumes.[3] These are usually only approachable through the keyword index which was introduced in 1963.[4] The index is printed at the back of each issue and is an alphabetic list of keyword phrases, with each abstract being indexed by, on average, five phrases. The phrases generally consist of three or four words, all significant, which are permuted so that they all appear as indexing points. For example: the entries for

Abstract 41444Q [Vol. 75 (6)], of which the title is "Transitions in Phases II-III-IV in high purity ammonium nitrate," are as follows:

> Ammonium Nitrate Phase Transition
> Nitrate Ammonium Phase Transition
> Phase Transition Ammonium Nitrate
> Transition Phase Ammonium Nitrate

Standardization in the keyword phrases is not achieved; the printing is small and constant use tends to strain the eyes. The delay in the issue of subject indexes at the time of the study mentioned above was averaging 18 months, which meant that a full search involved scanning the keyword indexes in some 78 issues. This task is so daunting that many people, particularly practising scientists, tend not to search the recent issues, thus missing the most up-to-date information.

For these reasons, therefore, ICI decided to provide cumulative indexes. The various alternatives considered during the development are discussed later. The result, however, is an index, based on the keyword phrases, arranged alphabetically, produced for 13 consecutive issues and stored on microfilm.

Each entry consists of:

Keyword Phrase (the first 80 characters)
Abstract Number
An Indicator for Patents
Title (the first 100 characters)
Specimen entries are shown in Figure 1.
The index is in upper/lower case format.

In addition to the keyword phrase index, a similarly formatted index based on patent assignees is produced. The subject index is produced on four spools of microfilm (equivalent to some 14,000 pages of computer printout), with a fifth spool holding the patent assignee index.

The *CA Condensates* records contain all the information needed to create the entries for each CA issue, and the index is produced by converting the weekly tapes into the index format. At the end of the quarter, these are

```
DTA PVC plasticized toluene resins
  74 06 023305V  Derivatographic studies on poly(vinyl chloride) plasticized with toluene-formaldehyde resins

DTPA amperometric titrn alk earths
  74 06 027741G  Amperometric complex-formation titration of traces of alkaline earths

DTPA bismuth detn
  74 06 0277480  Analytical applications of chelons.  LVII.  Chelometric determination of bismuth with DTPA (diethyle

DTPA cobalt detn
  74 02 009283E  Analytical application of chelating agents.  LIV.  Photometric determination of cobalt with DTPA (N,

DTPA potentiometry lead detn
  74 06 027761P  Potentiometric determination of lead with DTPA (diethylenetriaminepentaacetic acid) in the presence

DTPA treatment holophosphate phosphors
  74 08 036205N P Improving the lumen output of holophosphate phosphors by treating the phosphors with diethylenetriam

Eagles Nest mineral resources
  74 06 024195W  Mineral resources of the Gore Range-Eagles Nest Primitive Area and vicinity, Summit and Eagle Count'

Echinococcus antigen chromatog diagnosis
  74 05 020197A  Echinococcus granulosus and E. multilocularis: soluble antigen fluorescent antibody test

Echinocystis seeds biosynthesis diterpenes
  74 07 028885U  Biosynthesis of gibberellin precursors and related diterpenes

Echinops desert soils
  74 01 003112B  Early nutrition of Echinops spinosissimus on white-dune and desert soils

Eclipta demethylwedelolactones
  74 03 009684M  Isolation of demethylwedelolactone and its glucoside from Eclipta alba

Eclipta formylterthienyl
  74 01 001021X  Occurrence of desmethylwedelolactone and 2-formyl-,alpha,-terthienyl in Eclipta alba and the fertile

Ectocarpus copper tolerance
  74 03 010468A  Copper tolerance in the marine fouling alga Ectocarpus siliculosus
```

Figure 1. Specimen entries from microfilm index

merged into a single file, which is put on to microfilm using a Computer Output Microfilm (COM) device at an external bureau. The film can be available within a short time of the end of the period to which it refers.

Various alternatives were considered during the design of the index. These chiefly relate to the method of selecting indexing points. The three alternatives here were:

Single Terms Derived From Keyword Phrase and Title. This would give deeper indexing than keyword phrases only, but more complex computer processing would be necessary to eliminate duplicate terms. The main disadvantage, however, is intellectual. Frequently used terms such as "PVC" are relatively meaningless in isolation, and certain terms such as "acetic" are only valid when in compound terms—e.g. "acetic acid." The work of identifying such terms so as to produce a meaningful index is too great in relation to the potential benefits.

Keyword Phrases Plus Single Terms Derived from the Title. The problems of frequently used and compound terms would be reduced, but not entirely eliminated, by this approach. However, it would still be necessary to check for duplicate terms appearing in the title. Finally, this solution was rejected because it was felt that a "mixed" index of keyword phrases and single terms would be a poor compromise, having none of the advantages of either system.

Keyword Phrases Only. Examination of the keyword index showed that the significant concepts of the title are usually represented in the initial word of the phrases, and there is, therefore, no additional benefit from processing the title. This is illustrated in the example given above. The phrases printed are limited to 80 characters; examination has shown no cases where this limitation would, in practice, affect the phrase.

Another significant decision was on the inclusion of titles. Here the main point in consideration was the effect on the size of the index and, consequently, on processing costs. Eventually, it was agreed that the benefits (the nonrelevance of an abstract can frequently be decided on the basis of the title) outweighed the additional costs. Truncation of the title to the first 100 characters was agreed to simplify the printing process; less than 25% of titles exceed this length and loss of significance by truncation is lower than this.

The number of items to be handled is an important factor in any work on CA. It was estimated that the index would fill 14,000 pages of computer printout. This con-

firms the need to use microfilm. Experience with *Chemical Abstracts* on microfilm in ICI has been good, and the acceptability of microforms is increasing. All the main information units in ICI have microfilm readers available. The ability to produce good quality images economically, with upper/lower case format, by use of COM confirmed that we should use microfilm for this index.

The index has proved to be effective in use. General experience indicates that the time saved in a CA search justifies the cost; searching for references to a keyword over 18 months requires the scanning of six reels of film in comparison with 78 hand-copy issues. Also, and possibly more important, people who previously tended not to search the recent issues methodically are now able to do so with comparative ease.

The inclusion of titles has been found useful in practice, as well as in theory. One unforeseen benefit is that, if a reader/printer is used, the abstract numbers do not have to be noted as a prelude to the scanning of the abstracts— a copy of the relevant numbers, with titles, can be obtained easily, simplifying the whole operation.

Additional copies of the index can be produced easily by copying the microfilm, and the index can be available relatively cheaply in many locations within the Company.

Since this paper was first prepared in 1971, the permutations of the keyword phrases have been removed from the *CA Condensates* tapes. This greatly reduced the effectiveness of the microfilm index as the first term in the phrases remaining cannot describe adequately the content of the document. The terms can be re-permuted automatically; however, the index is no longer produced because of the cost of doing this satisfactorily. Nevertheless the underlying principle has been retained and microfilm indexes have been produced or are being considered for other collections. The production of the CA index again, in cooperation with other organizations, is still possible.

LITERATURE CITED

(1) Baker, D. B., "World's Chemical Literature Continues to Expand," *Chem. Eng. News* 49(28), 37–40 (1971).

(2) Brown, J. E., "The CAN/SDI Project," *Special Libraries* 60(8), 501–9 (1969).

(3) Skolnik, H., "Chemical Indexing: Management's Point of View," *J. Chem. Doc.* 1(1), 57–61 (1961).

(4) American Chemical Society, "Annual Report, 1963, Chemical Abstracts Service," *Chem. Eng. News* 42(10), 66–7 (1964).