The Adaptation to Computer Processing of Machine-Sorted Punched Cards Used for Retrieving Chemical Literature References*

STANLEY KIRSCHNER
Department of Chemistry, Wayne State University, Detroit, Michigan 48202

Received June 13, 1966

A system of coding, storing, and retrieving chemical literature references using machine-sorted punched cards has been revised and adapted for computer processing, in order to increase the flexibility and speed of the system. A new card format and a computer program have been developed which reduce to a minimum the amount of time a scientist must spend with the literature system. Provision is made for the complete printout of the authors, title, literature reference, and subject keywords of articles dealing with the given subject for which a search is conducted.

In 1957 the author (*1*) developed a system of coding, storing, and retrieving chemical literature references which is based upon the use of machine-sorted punched cards. In 1961 the system was modified slightly (*2*) so that it could be adapted to hand as well as machine sorting procedures, thereby making it somewhat more flexible than before. Although the original system was developed for the author's personal use, it has found use by other scientists at his own institution and has also been adopted by at least one industrial library.

A card was designed for the system which utilized IBM equipment and which provided for the punching of the author's name, journal reference, atomic numbers of the elements mentioned in an article, and several "subject code names," which referred to the subjects of interest to an individual scientist.

With the advent of the widespread availability of computers, Cameron (*3*) has suggested that it would be desirable to revise the system so that it would be applicable to computer processing. He did, in fact, make specific proposals for such modification to suit problems applicable to his own work.

DISCUSSION

The Revised Card. As a result of renewed interest in this field, the author has revised the format of the punched card (IBM-B16727, Figure 1), essentially changing only the field relating to the subject code numbers and atomic numbers. The remainder of the card is substantially unchanged. The old card, however, may still be used with the new system. One disadvantage of the old system (using subject code and atomic *numbers*) was that these numbers could not be interpreted on the card, because there would often be more than one number in a single column of a card. The new card provides for the use of two-letter "subject code" designations for each subject—a modification of the proposal of Cameron. There is space on the new card for 18 such "subject code letter pairs," (in columns 45 to 80), in addition to all the other information regarding author, reference, abstract, etc., which appeared on previous versions of this card. Lines still appear on the face of the card, and they are strategically placed so that information (such as an abstract) which is written or typed directly on these lines will not be mutilated by the punched holes. However, improvements in the system and the use of computers practically eliminate the need for writing on the cards at all.

Columns 1 to 5 of the new card contain space for a letter and a four-digit number to identify the card; columns 6 to 15 contain the senior (or first) author's initials and the first eight letters of his surname; columns 16 to 19 contain a four-letter CODEN abbreviation (*4*) of the journal name—*e.g.*, JACS = *J. Am. Chem. Soc.*; and columns 20 to 29 contain the volume number, page number, and the last two digits of the year of publication. Columns 30 to 44 contain similar information for the abstract reference, even including "L," the letter representing the portion of the column in which the abstract is to be found in *Chemical Abstracts*.

The Subject Code Letter Pairs. In place of a code number for a subject—*e.g.*, 50.1 (column 50, punch 1) to represent the subject "Coordination Chemistry;" 48.3 to represent "Stereoisomerism"—a unique two-letter combination is used to represent each subject—*e.g.*, CC to represent "Coordination Chemistry;" ST to represent "Stereoisomerism," Table I. There are 676 (that is, 26^2) possible unique code designations of this kind. At present, the author's list contains 105 subjects and subject code designations (plus 103 subject code designations for the chemical elements, using "Q" as the first letter for the representation of single letter elements: QC = Carbon), with the addition of about five subjects per year. Thus, having 676 designations available is more than ample for this author's system. Should more be needed, it would

* Presented before the Division of Chemical Literature, 151st National Meeting of the American Chemical Society, Pittsburgh, Pa., March 23, 1966.

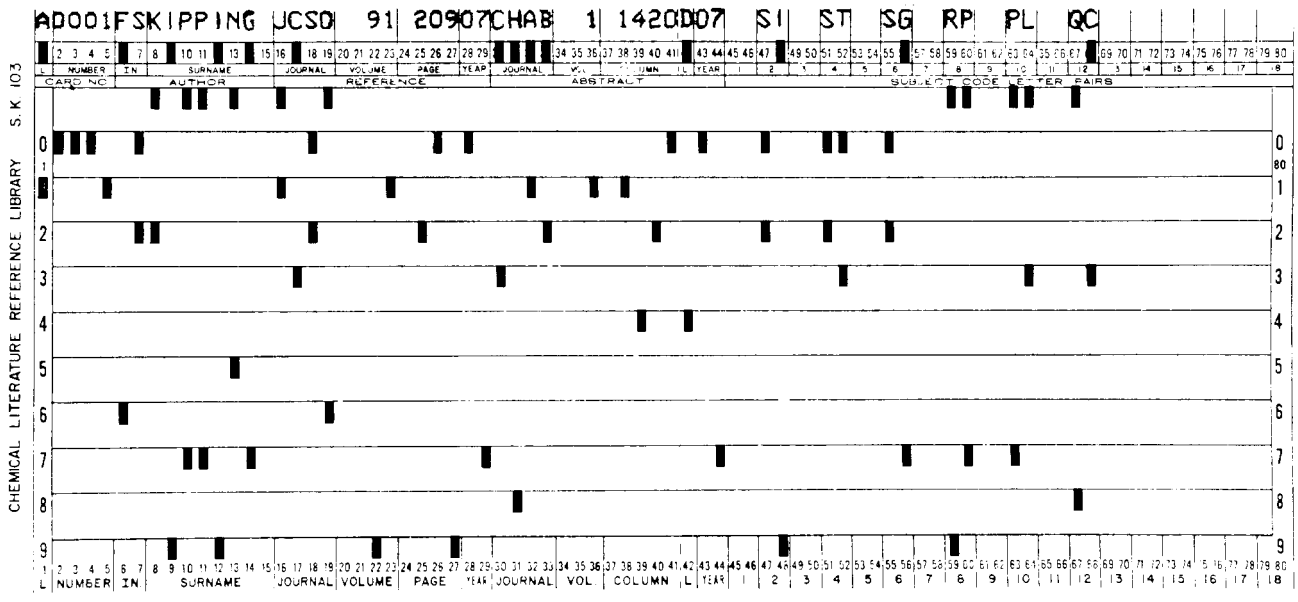

Figure 1. The revised card (IBM B16727) showing printing and punches for a typical primary reference card.

be possible to formulate a three-letter code (Cameron's suggestion), which would allow for 17,576 unique designations. Further, this system has the advantage of permitting the printing and/or interpreting of the subject code letter pairs directly on the card itself, along with all of the other information. Consequently, the entire card is now interpretable and usable for computer operation.

Provision is made for a maximum of 18 such subject code letter pairs on the primary card. To date, this has proved sufficient in the sense that there have not been more than 18 subjects in one reference (including the names of the important elements) which the author desired to code, and for which he may desire to initiate a search in the future. Should additional subject code letter terms be necessary, supplementary cards may be used to code and store them, provided such cards are also given the same identification number appearing on the primary card.

Table I. Part of a "Dictionary" of Subject Code Letter Pairs and Their Definitions

ABR.	SUBJECT INDEX TERM
AF	ABSOLUTE CONFIGURATION
AA	ACETYLACETONATES
AC	ACTINIUM
AB	ACIDS BASES
AL	ALUMINUM
AM	AMERICIUM
AC	AMINO ACIDS
AN	ANALYTICAL PROCEDURES
SB	ANTIMONY
AV	ANTIVIRAL COMPOUNDS
AE	APPARATUS
AR	ARGON
AS	ARSENIC
AO	ASSOCIATION (IONIC)
AT	ASTATINE
AP	ATOMIC ABSORPTION SPECTROSC.
BA	BARIUM
BK	BERKELIUM
BE	BERYLLIUM
BL	BIDENTATE LIGANDS
BG	BIOGRAPHY
BC	BIOCHEMISTRY
BI	BISMUTH
QB	BORON
BR	BROMINE

Supplementary Coding and Storage Cards. In addition to coding the information mentioned above onto the primary card, the author has found it desirable to utilize three supplementary cards: (a) an author card (or cards), which lists the initials and complete surnames of all authors associated with the article; (b) a reference card, which gives the exact reference in *Chemical Abstracts* terminology; and (c) a title card, which gives the exact title of the article. All of this information can be coded onto appropriate key punch forms by a person who is not technically trained.

Computer Program. A FORTRAN computer program (used on an IBM 7074 computer) was devised by the author which provides for the automatic printing out of all of the information on the primary and supplementary cards—*i.e.*, primary journal reference, abstract reference, all authors, and title of article. Further, this program permits the computer to accept an individual scientist's own "dictionary" (subject code definitions, Table I) and then to translate and print the subject code letter pairs into the complete words given in the "dictionary." The complete FORTRAN program for this (available on request from the author) requires only that the scientist indicate how many and which subject code letter pairs he desires in the references he wishes to locate. For example, Figure 2 shows a printout of a search through many cards for those articles having the four subject code letter pairs ST, SI, PL, QC (meaning stereoisomerism, silicon, laboratory preparations, and carbon, respectively). The computer searched for the primary cards having all four of these subject code letter pairs somewhere in columns 45 to 80, and, upon finding them, printed them and all the supplementary cards associated with them (author,

```
                                CHEMICAL LITERATURE REFERENCE SEARCH

NUMBER OF SEARCH TERMS AND SEARCH TERMS ARE  4 ST  SI  PL  QC
          STEREOISOMERISM
          SILICON
          PREPARATIONS-LABORATORY
          CARBON


NUMBER OF DEFINITION CARDS IS 207

    NO.  AUTHOR    JOUR VOL PG YR ABS VOL CCL YR    SUBJECT CODE LETTERS
   A0001FSKIPPING JOSC  91 20907CHAB  1 1420D07  SI  ST  SG  RP  PL  QC
   KIPPING FS                                                 A0CC1       C11
   J. CHEM. SOC., 91, 209 (1907)                              A0CC1       C12
ORGANIC DERIVATIVES OF SILICON                              A0CC1       C13   JQ

        KEY WORD INDEX TERMS
           SILICON
           STEREOISOMERISM
           ORGANOSILICON CHEMISTRY
           RESOLUTION PROCEDURES
           PREPARATIONS-LABORATORY
           CARBON


    NO.  AUTHOR    JOUR VOL PG YR ABS VOL CCL YR    SUBJECT CODE LETTERS
   A0008S KIRSCHNEJACS  80 75358                CC  SI  SG  RP  ST  QC  MT  PL  RTUC
   DHAR SK              DORON V              KIRSCHNER S           A0CC8       C11
   J. AM. CHEM. SOC., 80, 753 (1958)                              A0CC8       C12
RESOLUTION OF A HEXACOVALENT SILICON(IV) COMPLEX              A0CC8       C13   JQ

        KEY WORD INDEX TERMS
           COORDINATION COMPOUNDS
           SILICON
           ORGANOSILICON CHEMISTRY
           RESOLUTION PROCEDURES
           STEREOISOMERISM
           CARBON
           MOLECULAR-ATOMIC STRUCTURE
           PREPARATIONS-LABORATORY
           REPRINT
           UNUSUAL COORDINATION NUMBERS
```

Figure 2. A typical printout of the information in two coded references containing all of four specific search terms requested in a sort.

reference, and title cards). Further, *all* of the subject code letter pairs appearing on each of the primary cards printed are translated into printed words, thereby producing, in effect, a kind of brief abstract of each article located. The information found for two references having all four of the desired search terms is given in Figure 2. Other FORTRAN programs have been developed which allow searches for various combinations of author, journal, year, and subject. These programs are available on request.

Machine and Hand Sorting Potential. The machine and hand sorting capability of these cards as described previously (*2*) is not diminished by this revision. For example, the same IBM cards can still be used by hand sorting techniques if they are kept in numerical order and if a small handbook is kept which contains the same number of pages as there are subject code letter pairs (*2*), one subject per page. The card having a particular subject code letter pair would have its number on the page for that subject. To sort for that subject would mean turning to the subject page in the handbook and manually retrieving the cards whose numbers appear on that page.

Dictionary Printout. A FORTRAN program (available from the author upon request) has been developed for the purpose of generating a printout of the dictionary of subject code letter pairs and their definitions (Table I). Since the symbols of the chemical elements are well-known, these can be omitted from the table in order to simplify it.

Coding Personnel. A most important aspect of this system is that it reduces to a minimum the amount of effort and time which an individual scientist must devote to coding, storing, and retrieving chemical literature references he wishes to retain. Now, when a scientist wishes to retain an article, he merely writes in the margin of that article between one and 18 subject code letter pairs which give the subjects represented in that article (including element symbols). The scientist need do nothing more. Relatively untrained personnel can code the rest of the primary card and all the supplementary cards, and can carry out desired retrievals in the future.

ACKNOWLEDGMENT

The author wishes to express his sincere appreciation to the National Science Foundation for a grant (NSF-GP-5399) which contributed significantly to the progress of this work, to Robert Monroe and John Mack of the Wayne State University Computation and Data Processing Center, and to the staff of the IBM Western Processing Center, Los Angeles, Calif., for their assistance in the development of the computer programs described herein.

LITERATURE CITED

(1) Kirschner, S., *J. Chem. Educ.* **34**, 403 (1957).
(2) *Ibid.*, **38**, 526 (1961).
(3) Cameron, B. F., Department of Chemistry, University of Ibadan, Nigeria, private communication, in press, *J. Chem. Educ.*
(4) Kuentzel, L. E., ed., "Coden for Periodical Titles," American Society for Testing Materials, Philadelphia, Pa. 1963.