Bibliography Preparation by Computer

F. R. LIPSETT and F. D. BLAIR
National Research Council of Canada, Ottawa 7, Ontario, Canada

Received November 20, 1967

A system for recording a bibliography on magnetic tape for processing in an IBM 360/50 computer is described. Approximately 900 references, mostly from journals, are added yearly. Bibliographic details for each reference are typed on a specially designed card; each reference is assigned a number and classified by an alpha-numeric subject code.

In common with many scientists, the authors' bibliographies were first kept on plain 3 × 5 inch index cards. This system soon reached a stage at which sorting and filing the cards became overwhelmingly tedious, and edge-punched cards were adopted (*1*). After several years, searches of the edge-punched cards also became increasingly time-consuming and inaccurate. The system described below, based on the ideas of Chemical Titles tapes and the ASTM Coden System (*2*) was then devised. The system was intended for use as a personal bibliography (as opposed to a large system designed for a large number of users) and used a computer to carry out filing and searching procedures. It was put into operation at the beginning of 1967, and it has proved very successful (*3*). Although designed for a particular field of research (the luminescence of organic solids) it should be applicable with little or no modification to many fields of research.

OPERATION OF THE BIBLIOGRAPHY

Outline. References to be filed in the bibliography are obtained from a number of sources, including a selection of journals, Chemical Titles tapes, Chemical Abstracts, and Physics Abstracts. They are typed on specially designed cards (described below) which include a reference number, Coden abbreviation, and subject classifications as well as the authors' names, title, and so on. The material on the typed cards is keypunched and verified. Additions are then made to a master tape which includes the references and other data. Whenever entries are made in the bibliography the existing master tape is replaced by a new one. At the same time new references are printed out in a format shown in Figure 1. Also the number of references in each subject and bibliography classification is totalled and printed out. A loose-leaf book is kept which contains a printout of all references in order of entry and of the latest tabulation of references in each subject. Typed cards, keypunched cards, and at least one old master tape are kept on hand in case of accident.

The system for searching for references is outlined in Figure 2. References in a particular subject or subjects, by a certain author, or in some other class may be desired. The appropriate information is keypunched and with a suitable program the references are obtained quickly and printed out in a desired form.

Reference Card. The reference card designed for the bibliography is shown in Figure 3. The spacing of the lines is arranged to match the spacing of a typewriter (six per inch) and the spacing of the numbered blocks is similarly arranged (six per inch). The front of the card was drawn twice full size and reduced photographically. The cards are produced on white card stock by offset printing. Their size is 5 × 8 inches for convenient storage. The front of the card is printed in two colors—light blue and black. The light blue portions serve as a guide to the person filling out the card, while the keypunch operator can ignore most of the blue portions and concentrate on punching the black portions. The back of the card contains detailed instructions for entering the reference.

The top right block in Figure 3 is for the user's reference and is not keypunched. The control (column 1 of Figure 3) at present denotes a new entry for inclusion on the master tape, or calls for editing a reference already recorded. Editing may be required if, for example, a new subject classification is desired, if a misspelling is discovered, or if an abstract reference is to be added. Only columns 1 through 6 and the material to be added or

```
C0638 ADOLPH J AND WILLIAMS DF
      TEMPERATURE DEPENDENCE OF SINGLET-TRIPLET INTERSYSTEM
      CROSSING IN ANTHRACENE CRYSTALS.
      CODEN JCPS-A-0046-4248                          COPY ON HAND
      SUBJECT- S06 S12 S13 S14
      BIBLIOG.- 09
      PAGES 4248-51     1967  ENG   CA67  27180 D
```

Figure 1. Examples of possible printouts of reference shown in Figure 3

Top; printout in master list form. Bottom; printout in journal form

ADOLPH J. AND WILLIAMS D.F., TEMPERATURE DEPENDENCE OF SINGLET-TRIPLET INTERSYSTEM CROSSING IN ANTHRACENE CRYSTALS. J. CHEM. PHYS. 46, 4248-51 (1967) (ENG). (C.A. 67: 27180 D.)

altered need be typed on the card. A number 1 is entered in columns 7 and 8 if a copy of an abstract of the reference or a copy of the reference is on file. Columns 9 through 24 contain the ASTM Coden for the reference (*2*). Columns 26 through 60 are used for subject classifications pertinent to the author's field of research. Examples of subject classifications are given in Table I. The bibliographic classifications (columns 62 through 80) were included to facilitate the preparation of an annual bibliography on energy transfer in polyacene solid solutions (*4*). The code term 09 indicates that this reference is to be included in the 1968 edition of the bibliography. The remaining information is conventional. Care must be taken to insert author's names and initials in a certain manner to facilitate preparation of author indexes by the computer.

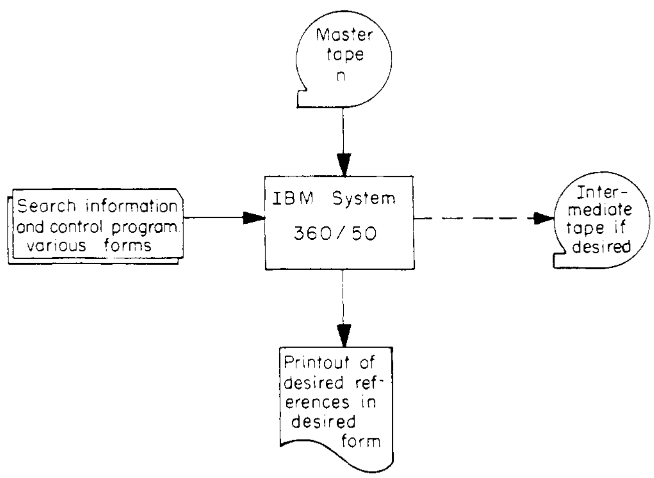

Figure 2. Diagram showing method of searching for references
Types of search and forms of printout are described in the text

Keypunching. Keypunching and verifying are done at the Computation Centre of the National Research Council. The operator is supplied with a few examples of keypunched references each time a batch of cards is to be punched.

Printouts. The references or information concerning the references may be printed out in various forms as follows:

MASTER LIST. Each time references are added to the master tape they are printed out in the form shown in the top of Figure 1. The references are arranged on numbered pages which are kept in a looseleaf binder.

TABULATION BY SUBJECT AND BIBLIOGRAPHY CLASSIFICATION. Each time references are added to the master tape the number of references in each subject and bibliography classification is counted and the totals are printed. This is a valuable aid when searches are to be made, and also indicates when a particular classification is "overloaded" or "under used" and should be revised.

AUTHOR INDEX. An alphabetical index using the first 20 letters (or letters and initials) of the author's name is prepared from time to time. This is an aid to checking whether a reference by a certain author is included in the bibliography.

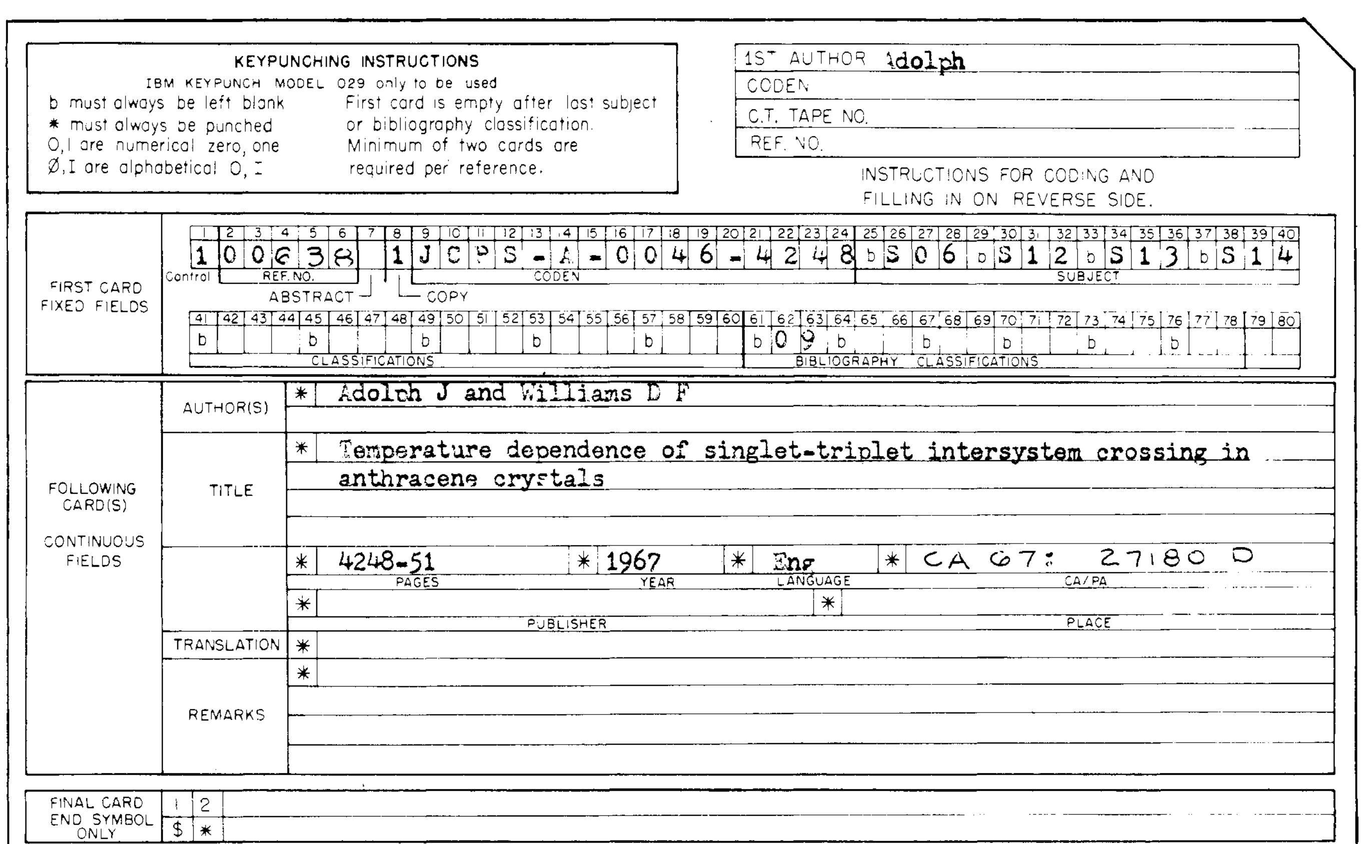

Figure 3. Reference card
Most of the card (including the top right block for the user's reference) is printed in light blue to give information and guide-lines for filling out. The block containing keypunching instructions; the spaces indicating first, following, and final cards; asterisks and dollar sign are printed in black. For a description see the text

It is also used in shortened form for preparing the author index for the energy transfer bibliography.

REFERENCE NUMBERS. The reference numbers of references in a particular classification or type of search may be printed out in a condensed format. This is primarily useful in preparing the energy transfer bibliography, when the references are arranged in classifications differing from those in the bibliography. Knowing the numbers of references to be included one can readily extract the typewritten cards from their file and arrange them in a new file.

REFERENCES IN A PARTICULAR SUBJECT CLASSIFICATION. The number of references in a particular subject classification may be found by consulting the tabulation described above. All references in the selected subject or group of subjects may be found by the appropriate search program and printed out in the desired form.

JOURNAL FORM. References may be printed out in a form suitable for publication. An example is given in the bottom of Figure 1. References printed in this form may, with suitable editing or instructions, be used directly by the printer or may be photographed. The Coden abbreviation has been used for the master list but the normal journal abbreviation (translated by the computer) has been used for the journal form of printout.

PRINTOUT BY NUMBER. When a group of references has been selected—for example, for a paper or for a section of a bibliography—they may be printed out in the desired order.

Table I. Examples of Subject Classifications[a]

T	=	Theories on Spectra
T01	=	Review articles and quantum theory
T02	=	Polyacenes—general
T03	=	Benzene
T04	=	Naphthalene
S	=	Experiments on Spectra
S01	=	Review articles
S06	=	Anthracene
S12	=	Triplet states and phosphorescence
S13	=	Radiationless transitions
S14	=	Delayed fluorescence
R	=	Related Subjects
R01	=	General articles on luminescence
R02	=	Molecular structure (incl. impurities)
R03	=	Decay times—fluorescence

[a] Complete system includes 175 classifications.

System Details. The computer used is an IBM System 360/50 located at the Computation Centre of the National Research Council. When the system was planned, three program languages were available—Fortran, PL1, and Assembly language. Fortran was not selected because it was felt that the version in use at that time would have restricted the flexibility of both the data processing and input-output formats. PL1 might have been a desirable possibility because of its flexibility and possible general application to other computer systems. However, Assembly language was finally chosen, as it allowed all of the System/360 facilities to be utilized to the fullest extent and allowed various data formats and subroutine structures to be tried with greater ease than would either of the other two languages. As the system is to be modified to utilize disk storage using the index sequential mode

Table II. Summary of Programs

Program	Function	Printer Output	Tape Output	Remarks
Main program	Sets up original tape. Adds references to tape. Checks data formats. Corrects existing references. Updates accounting information.	Card images. Gives updated accounting information only.	Master tape only.	Card input. Variable length tape records.
Search program	Searches for individual references by number. Scans over a range of reference numbers. Searches for references by a single author. Completes accounting information.	Master list form. Journal form. Accounting information (complete).	None.	Will accept either master tape format or special group classification format.
Secondary search program	Searches by subject or bibliography classification, or by Coden.	Master list form. Reference numbers.	Master tape format. Special group classifications.	Allows assembly of references in special groups for resorting.
Author index	Searches for all authors and sorts into alphabetic order. Carries along reference numbers associated with authors.	Authors' names sorted alphabetically by first 20 characters. Printed in groups according to first letter of surname. Reference numbers included for each author.	None.	Limits may be placed on the search such that: *a*. Only a given range of reference numbers will be searched. *b*. Only a given range of first letters will be searched. *c*. Combination of *a*. and *b*.

of operation, very little difficulty should be encountered in rewriting the programs in PL1 using the original flow charts.

Table II contains a program summary outlining the functions of each program. The programs were set up in the order given to allow for editing at various points in the processing sequence. The following should be noted:

Forms of printout other than the four described above may be added without disturbing other parts of the system.

Intermediate output tapes with various formats have been provided to allow for multiple levels of searching. In this way the computer may be used to search for references included in combinations of subject or bibliography classifications, author's names, Codens, and so on. The references thus found may be stored on the intermediate tape and then resorted by the computer in a particular order for output. For example, the master tape may be searched for references for a bibliography (with a coded classification number) which may be stored on an intermediate tape. These references may then be sorted into subject classifications, arranged alphabetically, and printed out in journal format.

The system may be used for disk storage without modification provided sequentially-organized data files are desired.

Further details of the programs may be obtained from the authors.

When the reference card was designed, one of the main aims was to make the bibliographer's job simple. From the viewpoint of the programs, it was desirable to keep storage to a minimum while retaining all information necessary to produce bibliographies. It was also desirable to make the format efficient for computer searches for specific types of information. These three objectives dictated dividing each reference into two parts, fixed format data and variable length data. This is illustrated in Figure 3.

The fixed data described above can be punched on a single card. The information for which the machine normally searches begins in predetermined columns, thus making the search routines efficient. Appropriate blanks have been left for the convenience of the typist and the keypunch operator. The Coden field may contain special key words, such as report, book, or thesis which will cause special formatting when the reference is printed in a bibliography.

In setting up the card for the variable length fields, it was felt that no restrictions should be placed on the length of the fields or on the number of authors' names to be allowed. To separate the various fields, asterisks were used as delimiters. Each item must begin with an asterisk and if the field does not contain an entry, the asterisk must still be included on the punched card. The cards (other than the fixed field card) are punched continuously without regard to where the end of the card physically breaks into the data. Each reference may extend over 25 cards. All of these cards are re-assembled in the computer into one continuous field. At this time, trailing and leading blanks are removed from each field for efficient use of storage, and the complete reference, including fixed and variable data, is outputted as one variable length record under control of the operating system's data management routines.

In addition to the two types of punched cards just described, a control card is required after each reference. This card denotes the end of the reference and contains a dollar sign and an asterisk ($ *) in columns one and two, respectively.

RESULTS AND REMARKS

At the time of writing, the bibliography included 910 references, which were enough to test its capabilities thoroughly. It has been found to operate even better than hoped. The master list of references and author index in book form are a convenience almost unobtainable by previous systems employing conventional cards. Searching has proved rapid and comprehensive, as some examples will illustrate. In one instance a search for nine references in a particular subject classification was completed, and the references printed out in master reference form, in 23 seconds of machine time. In another instance the 249 numbers of all references in bibliography classification 09 were printed out, and the total checked, in 74 seconds of machine time. The author index (which includes all authors) for the 910 references was prepared in 4 minutes of machine time. The system has been adopted in modified form by a group in another part of the National Research Council and may be extended in some way as a general facility.

Even though the system works very well as described, a number of improvements are planned or envisaged which should make it more efficient and useful. In particular, as mentioned above, it is planned to store the bibliography on a Disk Storage Unit rather than on magnetic tape. The Disk should be more efficient for the present application because it gives random access as well as sequential access. Also, if remote terminals were installed for time-shared operation of the computer, this would reduce the turn-around time for most searches.

LITERATURE CITED

(1) Hoyle, W. G., *Eng. J. (Eng. Inst. of Canada)* **41**, 61 (1958).
(2) *Coden for Periodical Titles, Vols. I–II.* ASTM Data Series DS 23A. American Society for Testing and Materials, 1916 Race St., Philadelphia, Pa. 19103 (1966).
(3) A similar, less elaborate system was recently described by C. N. Gillis, J. CHEM. DOC. **7**, 98 (1967).
(4) Lipsett, F. R., *Mol. Cryst.* **3**, 1 (1967).