THIS EDITION

September is traditionally the first month of Spring, the season of poets, pastoral melodies, garlands of flowers - and romance. We in the Computer Centre refuse to accept the age-old cleavage between the arts and the sciences, and present to the reader, in this issue, our impression of the Spring activities (and celebrations) of our two computers!

To return from the ozone-blue heights of Spring, the Bulletin also brings to light a case-study of the Computer and its contribution to library work, described by Mr. John Row working in conjunction with the Thatcher Library. Other items of news include the Australian Computer Conference in Adelaide, and a clarification of the new arrangements for clients wishing to consult programmers.

STAFF OF THE COMPUTER CENTRE

In this month's Bulletin, we farewell two members of staff, Leonard Mor and Mrs. Dianne Neill.

Len, who has been with the Department since April this year as a Senior Demonstrator, has left for England. We wish him "Bon Voyage" and a happy stay.

Dianne was one of the new Machine Operators who were appointed in May. To Dianne, go our sincere wishes for much future happiness.
ADELAIDE COMPUTER CONFERENCE

The chief topic of conversation among both staff and students last month was the Australian Computer Conference held in Adelaide from August 11th - 15th. This was the fourth conference convened in Australia, the first three at Sydney (1960), Melbourne (1963), and Canberra (1966) being sponsored by the Australian National Committee on Computation and Automatic Control.

The Adelaide Conference offered three broad streams of discussion: scientific, technical and commercial. The commercial aspect, in particular, was stressed, each day concluding with a panel discussion of questions arising from the papers delivered. Broadly, the commercial, industrial and management stream researched management information systems, data processing administration, and the application of scientific methods in commerce and industry. The technological side encompassed topics such as hardware, software, and computer languages, while the scientific and professional stream discussed subjects including simulation, contour mapping, and medical, educational and engineering applications. There was, however, a considerable amount of overlapping and cross-referencing which tended to reduce the effectiveness of these three divisions.

Films were screened on various applications of computing, and several displays of computer components and books were offered for general interest.

The Department of Computer Science was ably represented by two speakers as well as a host of student spectators. John Noad, Lecturer in the Diploma in Information Processing, presented a paper entitled *A File Organisation for Single Disk Drive Systems*. On behalf of Ian Oliver, formerly a Lecturer in the Diploma in Automatic Computing, Dick Kelly presented a paper *Command, File, and Task Structure for a Medium-sized Time-shared System*.

The fifth Computer Conference will be held in Brisbane in 1972. We have merely three years to prepare for an influx of perhaps 2,000 people who must be wined, dined, accommodated, assembled together for lectures and discussions with ease, comfort, and no hint of over-crowding. In the absence of better facilities, we suggest the hire of an air-craft carrier.
DON'T BOTHER US NOW
IT'S SPRING
CONSULTATION WITH PROGRAMMERS

Because of the extreme shortage of staff this year, it has been difficult for programmers to set aside specific periods for consultation with clients about their programming problems. As a result, a somewhat confused arrangement has existed for most of the year.

To avoid inconvenience to both clients and programmers, we ask that clients seeking advice first contact the Administrative Officer (Mr. John Jauncey, extension 8471). He will then arrange an appointment with a programmer at a time suitable to both parties.

We hope that this procedure will prevent awkward interruptions, and give the maximum amount of assistance to clients.

PDP 10 FORTRAN IV ERRORS

1. Checking of Array Subscripts

In PDP 10 FORTRAN IV, no check is made of the value of any subscript at execution time. The same situation applies also to the GE 225, but it is interesting to note that in CARD FORTRAN, checking is done for any array reference using more than two subscripts.

In PDP 10 FORTRAN, the use of an out-of-range value may result in the depositing of data outside the addressed array, or, even worse, over program locations. Such an error may not be detected until later in the program where it may manifest itself through improbable results, or as an illegal instruction if the data were deposited in a program location.

To summarize briefly, be on the guard for the three most likely ways in which such an error will be revealed, viz:
(a) an illegal memory reference
(b) an illegal instruction or UUO
(c) invalid results

2. The use of a function in a Double Precision or Complex array subscript may result in the function name replacing that of a mathematical utility routine such as CFDV used later in the expression. For example, incorrect coding is generated by the following:

```
DOUBLE PRECISION D(5)
R = 1/D (IFIX(R))
```

Thus, never use a function in the subscript of a Double Precision or Complex array.

The problem can be avoided quite simply. The following coding would be acceptable for the above example:

```
I = R
R = 1/D(I)
```

3. Octal constants greater than $2^{35}$ give a constant overflow message at compilation time when defined in assignment statements.

For example:

```
B = "7777700000"
```

The situation can be remedied by using the DATA statement, i.e.,

```
DATA B/"7777700000"
```
4. A literal constant composed solely of two adjacent single quotes, generates no error and refers to the previous constant generated.

For example:

\[
A = 'A' \\
B = ''
\]

It is advisable to take note of this feature.

Users are reminded to report any suspected errors to the Administrative Officer (Mr. John Jauncey, extension 8471).

PROJECT THATCAT

J. Row

BACKGROUND

This article describes a computer cataloguing project undertaken by Mrs. J. Guyatt of the Thatcher Memorial Library. The system analysis and programming were done by the author on a part-time basis while studying towards the Diploma in Automatic Computing. Mrs. Guyatt has described the Library aspects of this work in an article to be published this month in the Australian Library Journal. This discussion will be confined to the programming and systems aspects of the work.

THE PROBLEM

The Thatcher Memorial Library provides a library service for the external students of the University. For a number of years, it has produced a comprehensive, annotated catalogue to aid external students in the selection of suitable books for study. The catalogue is intended to substitute, as far as possible, for the
opportunity internal students have of browsing among the shelves of the Library. To accomplish this purpose, the catalogue has three main sections:

(1) **AUTHOR LIST** has an entry for each book with the following information:

<table>
<thead>
<tr>
<th>AUTHOR</th>
<th>TITLE</th>
<th>DEWEY NUMBER</th>
</tr>
</thead>
</table>

(2) **DEWEY LIST** has an entry for each book:

<table>
<thead>
<tr>
<th>DEWEY NO.</th>
<th>AUTHOR</th>
<th>TITLE</th>
<th>DATE OF PUBLICATION</th>
<th>NUMBER OF PAGES</th>
<th>ANNOTATION</th>
</tr>
</thead>
</table>

(3) **SUBJECT LIST.** Under each subject is a list of DEWEY NUMBERS of relevant books. Subjects can be sub-divided where applicable, and a particular book may be included under a number of subject headings.

A two volume edition was completed in 1965. Since then, supplements have been produced on three occasions. This means that four separate lists might be examined before a particular book is found. The Library staff were then faced with the formidable task of accumulating approximately 25,000 titles, by hand, in the traditional manner:-

1. Sort shelf lists under Dewey Number
2. Type the Dewey List from the shelf cards
3. Sort shelf lists by Author
4. Type the Author List from the shelf cards
5. Sort the shelf list by Subject, one subject at a time, type, sort the next subject, type, etc. (One book may appear under a number of subjects).

Mrs. Guyatt decided to investigate Computer methods.

**PROBLEM DEFINITION**

At first sight, the problem appeared to be a straight-forward one that would be adequately handled using GECOM and FOR ward-SORT-MERGE Programs on the GE 225. Unfortunately, closer examination revealed a number of difficulties, viz:
1. Most of the information (including keys for sorting) could not be restricted to fixed fields.

2. Field lengths vary dramatically. For example, a particular book might have an annotation 20 lines long (=2000 characters), or no annotation at all! Most of the relevant fields vary in a similar way. In one notable example, the title of a book extended to 1000 characters, with an annotation a mere 20 characters. (The book itself was only 12 pages.)

3. Difficulty in automating complex library filing (sorting) rules. For example, we are all familiar with a Dewey number of the form

   322.4 PAR

but the following Dewey numbers are also legitimate, and must be sorted correctly to the last level of significance:

   332.6730994 FIT
   810 DICKE(E) 3A BIN
   364.1430973 UNI WAS
   370.10973 DEW 3 WIR

The first (numeric) field of the Dewey number (i.e. the 322.4 part) may also have an alphabetic suffix. These must be sorted in the following way:

   820
   820.9
   829.1
   820A
   820.9A
   820NZ
   820.93NZ

Characters such as "(" , ")", ",", "-", have special significance to the computer, and therefore must be eliminated from the sorting key. Clearly, special keys need to be generated in order to obey filing rules. The generation of these keys inevitably requires basic character manipulation at a machine instruction level.
Library staff were adamant that, if the computer could not learn to perform all these tricks, it was no use to them.

From this initial study evolved a set of considerations for the system design:

1. For all purposes, fields should be completely free on input and output and have no practical limit on size.
2. Because of the large amount of information involved, efficiency of input, output, and tape storage is very important.
3. Standard library rules for filing should be obeyed as closely as possible.
4. The Input program should perform extensive error checking and produce meaningful diagnostics.
5. Because of the limitations on character set, attention should be paid to the layout of information on printout, for maximum legibility. The output from the High Speed Printer should be suitable for photographic reduction and direct printing without an intermediate stage.
6. Sorting and Merging should be done by programs generated by the FORWARD-SORT-MERGE package. Sorts should be tape to tape, with minimum complication.
7. Programs should be written in the Assembly Language of the GE 225 (GAP) for maximum flexibility and efficiency.
8. All required character manipulation should be done in the input and output programs which are certain to be bound by slow input/output devices.

DESIGN CONSIDERATIONS

The requirements of the system can be summarized as follows:

1. Produce annually a cumulative catalogue containing the three sections

   AUTHOR
   DEWEY (ANNOTATED)
   SUBJECT
2. At intervals during the year, produce an accession list of all books received in the interval between the last accession list or catalogue (probably every 3 months). The accession list should have only the Dewey section.

3. The process of preparing information for the computer should be part of the day-to-day operations of the Library; it should not necessitate a concerted effort by a number of staff.

The three sections of the catalogue are maintained as three separate tape files which are multi-reel, although initially, a single reel for each file will be sufficient.

The system has four broad phases, (see Figure 1):

1. Input
2. Sorting
3. Merging with Master File
4. Output

When the library acquires a book, a shelf card is created, containing all the information required for the catalogue. At certain intervals, these shelf cards are sent to the Computer Centre for punching. The information is punched directly from the shelf cards onto computer cards in accordance with the card input format.

When an accession list is required, the cards are run with the input program to produce an unsorted Dewey tape. This phase is usually repeated several times until all errors have been eliminated. The tape is then sorted and run with the Dewey Output Program to produce the Accession List. The tape is not stored for future use. Tapes and handlers are allocated so that this operation can proceed from card input to printing in one run. Typically for 2000 cards (about 400 books), this would take about 14 minutes.

At the end of the year, all the cards are run with the input program to produce Dewey, Author and Subject transaction tapes, which, after sorting, can be
merged with the master files. The Output programs then produce the catalogue from the updated master files. Only one Input run should be required at this stage because the cards have been "debugged" and checked on a previous occasion.

Broadly speaking, the design of the system depends very much on the design of the tape formats. A great deal of the work of formatting the catalogue is done by the output programs. The tape formats are designed to increase efficiency in sorting and storage and to preserve the integrity of the information for ease in updating.

INPUT PROGRAM

Here is an example of the cards punched for a particular book:

$370.973 KAR/KARIER, C.J./MAN, SOCIETY AND EDUCATION/1967/334/ B763
A HISTORY OF EDUCATION FOUCSED ON THE CONCEPTS OF HUMAN NATURE AND COMMUNIT Y WITH CHAPTERS ON A FASCIST AND COMMUNIST VIEW OF THE FUNCTION OF THE AMER B764
ICAN SCHOOL AND THE UNITED STATES SUPREME COURT AND EDUCATION./ B765
EDUCATION/U.S./HISTORY//EDUCATION/PHILOSOPHY/// B767

1. Fields are delimited by /

2. The three main sections of the input are:

   DEWEY NO/AUTHOR/TITLE/DATE/PAGES/
   ANNOTATION/
   SUBJECT ENTRY//SUBJECT ENTRY///

3. The sequence number is used by the input program only as a check of the integrity of cards belonging to a particular book, (e.g. a missing card will produce an error).

4. The $ at the beginning of a book gives a certain amount of redundant checking and allows recovery in case of an error.

5. Words can be split from one card to the next.

6. A field may extend over any number of cards up to 30.
7. There may be any number of subject entries, each delimited by //. (The single / within a subject entry indicates a sub-heading.)

8. There can be up to 2 sub-headings for a subject.

9. Because subject entries are used directly as keys, they are restricted to 45 characters.

The Input Program checks the format of input and rejects any book which has errors. A listing of all input data and error messages, where applicable, is produced. There are 35 different error messages. (See Figure 2 for an example of the input listing.) The special sorting keys are generated and tape records written.


The records are of fixed length (41 word) to simplify sorting.

<table>
<thead>
<tr>
<th>SORTING KEY</th>
<th>BOOK SEQ.NO.</th>
<th>CARD IMAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Text information is stored on tape in the same format as input cards. The sequence number ensures that the records belonging to a particular book will remain in the correct sequence after the sort. It is generated by the Input Program and bears no relationship to the card sequence number. A book may occupy as many records as are needed.

2. Author:

Same format as Dewey.

3. Subject

A record is written on the subject tape for each subject entry for each book.

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>DEWEY KEY</th>
<th>DEWEY NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The subject tape must be sorted under subject first, then Dewey number. It is therefore necessary to include the Dewey sorting key. The card image of the Dewey number (with ",", ",", ",", etc.) is also included for final output in the catalogue. All interpretation of subject lists is done by the Output program.

At the end of an input run, a summary of the run is printed.

SORT & MERGE

The SORT and MERGE programs used were generated using the FORWARD SORT MERGE package on the GE 225. They include Input, Output, and Squeeze Coding Elements to check for errors and effect deletions. Sorting times are quite fast. For example, 1200 books sorted in 6 minutes for a Dewey list.

OUTPUT PROGRAMS

The Output Programs do a large amount of the work in formatting the catalogue.

Compacted information from tape is unpacked and reformed to improve readability. Page numbers are also printed. Input from tape is considered to be a string of characters rather than a series of card images. This allows the program to select line lengths, spacings and indentations consistent with readability. Words can be split across records on the tape, but this is not allowed on the printed output. [For each line of output, a line length is found so that the last word is complete. Blanks are then skipped so that the next line is left justified.] In general, the format of the printed page bears no relation to the format on the cards at input.

Available page space is also checked so that the entry for a book is not split over two pages. If there is insufficient room for an entry, a new page is started. (For examples of the Author, Dewey and Subject listings produced, refer to Figures 3, 4 and 5 respectively.)

In particular, the Subject Output Program does a lot of work interpreting the Subject file. Sub-headings are placed on a new line, indented depending on the level and preceded by a "-". If more than one book appears under a subject heading, the heading is not repeated. If more than 10 books
appear under a heading, the program looks for a common term in the Dewey number so that a general number covering a group of books can be inserted. Books are grouped under main headings and sub-headings as shown in Figure 5. A broad outline of the unpacking algorithm is given in Figure 6. Use is made of a recursive formatting routine.

CURRENT STATUS

The only part of the system which has not been used at present is the merging section. A catalogue was produced from the books received in 1968. Two accession lists have been produced from the 1969 books, and these will be merged with the 1968 tapes at the end of the year.
Figure 1. System Flow-Chart

- Cards
- Input
- Phase
- Subject
- Input Program
- Dewey
- Listing
  - Three tapes written
  - Errors etc.
- Author
- Any sorted tape
- Output program
- Catalogue
- New master
- Merge program
- Old master
- Similar process for Author, Dewey & Subject files
Figure 2. Example of the Listing Produced by the Input Program
Figure 5. Example of Subject Listing
Figure 6. Subject Unpacking Algorithm