COMPUTERISED DATA PROCESSING FOR BRITISH TECHNOLOGY INDEX

E. J. COATES

Indexing of all kinds inevitably calls for a certain amount of clerical work and, in the case of a periodically published index working to a dateline, careful apportionment of time between the indexing process proper and the necessary clerical tasks is a necessity. British Technology Index adopted computer assistance in the performance of clerical work in April 1968, after a research programme which dates back to 1964.* The motive for adopting computer assistance in this case was the need to maintain and if possible improve the currency of the index, which attempts to record articles from British technical journals within seven weeks of their original publication. Under the former manual regime indexing had to slow down at a point about two-thirds of the way through each month in order that the related clerical work could be accomplished in time. So much in the final publication depended upon the accurate carrying out of this work that it could not be allocated to clerks. It was in fact part of the indexers' work load.

Each month's issue of the British Technology Index comprises 2,000 to 2,500 entries, plus 5,000 to 6,000 cross-references. The cross-references are generated or initiated by the subject index headings which are attached to entries. That is to say, if a particular subject heading is used in a given issue of the Index it calls up a particular set of cross-references which it will be necessary to include in the issue. Some of the cross-reference copy is used repeatedly and is kept in store until required; previously unused headings demand the production of fresh copy for the corresponding cross-references, and this was done concurrently with indexing. At the point mentioned above two-thirds of the way through the month, the file of entries and headings needed to be checked against the store of cross-references, and the necessary cross-reference copy withdrawn from store for the printer. This task of extracting 5,000 to 6,000 cross-references, including that of extracting cross-references to cross-references, took up many man-hours, and in minor emergencies such as staff sickness could bring the indexing process to a halt altogether.

It is this task of producing cross-reference copy from a given set of subject headings which has now been assigned to the computer. The complete process takes a little over an hour each month, and there is, in consequence of this saving, an increase of available time for indexing. A by-product of the main process is a monthly updated authority file of indexing decisions produced.

* A feasibility study was undertaken by the University of Newcastle-upon-Tyne Computing Laboratory in 1966-67, with the support of the Office for Scientific and Technical Information. The University of Newcastle-upon-Tyne Computing Laboratory also undertake for BTI the production data processing described in this article.
by the computer in print-out form. An authority file is of course needed continuously by indexers to check tentative headings and so ensure consistency. The manual authority file which the computer print-out forms will eventually supersede is in index-card form and up to April 1968 had accumulated nearly 250,000 records. Its maintenance was an increasingly heavy burden upon indexers’ time, and it was becoming steadily more difficult to consult. Also there could be only one copy of it, whereas if it is found desirable each indexer can have his own copy of the print-out authority file. In the specimen reproduced the asterisked items are headings; lines which start at the second indentation from the left are cross-references, and those at the third indentation are reversed cross-references for synonyms and related terms. The latter do not appear in the final product.

To leave the by-product and return to a consideration of the main product, namely cross-reference copy, the general principle is that the editorial office punches coded subject headings to paper tape by means of a tape-output typewriter. On receiving these, the computer produces the appropriate cross-reference copy corresponding to the subject headings. The computer achieves this partly by extracting cross-references from a magnetic-tape store, which is closely analogous to the former manual store and partly by manipulating the component words in a subject heading according to prescriptions in the programme. The first process of extracting from magnetic-tape store, is easy to understand. The computer transfers the subject heading data to magnetic tape from the paper tape supplied and proceeds to sort it. It then compares this sorted file with another sorted file of coupled terms constituting the store of ‘related heading’ and synonym cross-references. When it discovers a term match between input subject heading and store terms, it copies the store terms concerned on to another magnetic tape and at the same time expands the information copied to the form required at the final output. An example will make this clear.

Let us suppose that part of the related heading and synonym store contains the following coupled terms:

- AIR COMPRESSORS/COMPRESSORS
- AIR CONDITIONING/BUILDING
- AIR CURTAINS/INSULATION, THERMAL
- AIR GAPS/INSULATION, ELECTRICAL
- AIR HEATING/HEATING
- AIR MOTORS/PNEUMATIC MACHINERY
- AIR POLLUTION/SANITARY ENGINEERING
- AIR SPEED INDICATORS/ AIRCRAFT, INSTRUMENTS

and let us suppose that part of the sorted file of the month’s input subject headings reads as follows:

- AGGREGATES
- AGRICULTURAL MACHINERY
- AIR HEATING
- AIR TRANSPORT
- AIR TURBINES

The computer systematically compares the two sets of terms until it discovers that there is an identity match at AIR HEATING. It then transfers AIR HEATING/HEATING to another tape, and performs a combined expansion and transposition operation on the terms to give the final output

\[ \text{HEATING related headings} \]

\[ \text{AIR HEATING} \]

It will be seen that under this system, BTI has to inform the store—by means of punched paper tape—when a subject heading is used for the first time and what the ‘related heading’ and synonym term is. Once the information is put into store it remains available for further use without any action by BTI beyond inputting the subject heading.

The second process—that of generating inversion ‘See’ cross-references, by computer manipulation of the component terms of the subject heading is much more complex and is probably an original concept in computer-assisted production of indexes. These manipulations were quite closely subject to rule under the manual system. It was necessary to refine and close any loopholes in the rules for the purpose of making a computer programme.

The simplest possible pattern of manipu-
lation is illustrated by the following example:

For input subject heading
  WATER Flow Pipes
Output cross-references
  PIPES Flow Water
  See WATER Flow Pipes
  FLOW WATER See WATER Flow

Computer package programmes which will perform a manipulation of this kind have been in existence for some time, and if this manipulation alone were all that is required, there would have been no problems. However, British Technology Index is designed primarily for the user whose interest is in highly specific subjects, and one corollary of this is that our cross-reference manipulation system must never result in loss of specificity in relation to a particular kind of thing or action. If we slightly change our example and employ the simple symmetrical pattern given above we have the following:

Input subject heading
  WATER Flow Turbulent Pipes
Output cross-references
  PIPES Turbulent Flow Water
  See WATER Flow Turbulent Pipes
  TURBULENT Flow Water
  See WATER Flow Turbulent
  FLOW Water See WATER Flow

The shortcoming in this result, from the point of view of a specific subject oriented index, is that at FLOW we have lost the idea of turbulent flow. It is quite true that the item is about Flow of water but the reference fails to say that it is about turbulent flow. In this case therefore where two consecutive terms in the subject heading string represent a Thing or Action (an Action in this case) followed by a term designating a Kind, then these are treated as inseparable and resort is made to permutation. Thus, the last cross-reference in the example above would read

  FLOW Turbulent Water
  See WATER Flow Turbulent

The computer can perform this variant manipulation if we can find some way of informing it that Flow Turbulent in the subject heading string is a special situation requiring this variant manipulation, unlike Water Flow or Flow Pipes or Turbulent Pipes. The special signal required is in fact put in by the indexer whenever he finds two consecutive terms in his subject heading exemplifying a Thing (or Action) plus Kind situation. The signal takes the form of differentiated punctuation in the input subject heading. Thus to produce the simple symmetrical pattern of cross-references given in the first example the input is punctuated with colons

  WATER : Flow, Turbulent : Pipes

To produce the superimposed permutation required whenever a term specifies a Kind of the concept represented by the preceding term in the subject heading, the punctuation in front of the 'Kind' term is changed to a comma, giving

  WATER : Flow, Turbulent : Pipes

The visual effect of this method of punctuation is to suggest that Flow and Turbulent are more tightly bound together than Water and Flow or Turbulent and Pipes, and indeed it is this relative inseparability which is reflected in the changed pattern of inversion references produced. One other thing has to be remembered by the computer by means of a special instruction in its programme. It reproduces the colons on the left hand side of the cross-references but not the comma, giving

  PIPES : Turbulent flow : Water See
  TURBULENT FLOW : Water See

but yet again where the order of the permuted terms is the same as that in the subject heading, the comma is retained

  FLOW, Turbulent : Water See

Frequently a type specifying term is not an adjective but the name of a material. In this case also the superimposed permutation pattern is used, but the signal is a semicolon, which is repeated in the output cross-references. For example, the flow of water in concrete pipes would be given as

  WATER : Flow : Pipes ; Concrete

The computer programme can also cope with the fact that on occasion two or more kind-specifying terms may appear consecutively in a subject heading string. The second kind-specifying term may relate to the first kind-specifying term or to the entity being specified. Two contrasting subject heading
inputs which show this effect are the following:

- PIPES ; Steel, Stainless
- PIPES ; Steel Ribbed

Stainless refers to steel, but ribbed refers to pipe.

In the examples given so far every term in the subject heading string initiates a cross-reference, but this is not always the case. In the example

- WATER, Saline : Conversion : Ion exchange
- WATER : Extraction : Rivers

we make no cross-references beginning respectively with Measurement, Conversion, or Extraction, because we do not think that any user will need to look up everything on such pervasive concepts. We therefore need to signal to the computer whether or not each term in a subject heading string initiates a cross-reference. This is done by means of a second punctuation character (or 'second operator') between words, which having served its purpose is deleted on final output. The second operator is inserted in the following manner:

- WATER :. Flow : Pipes

The full stop before Flow and Pipes instructs the computer that each of these words initiates a cross-reference.

In the example

- WATER : [Extraction : Rivers

The bracket before extraction instructs the computer to make no cross-reference commencing with that term.

We have a list of about 40 terms which are so diffuse in their application that it is normally desirable that they should not appear either in the leading or second position in a cross-reference. These have a special code signal which conveys to the computer both an instruction that they do not initiate a cross-reference and a direction which 'relegates' them one place to the right in the inverted cross-reference string. This is illustrated by

- WATER :. Flow : Measurement :. Weirs

which gives on output cross-references beginning

- WEIRS : Flow measurement : Water. See WATER : Flow measurement
- WATER : Flow : measurement : weirs
- FLOW : Water. See WATER : Flow

Another second operator provides the facility for treating any two subject heading elements as one for manipulation purposes, wherever this may be necessary. The converse, the splitting up of a phrase forming part of a subject heading is also provided for in the programme. In all, the inversion reference programme comprises 28 separate instructions, and in practice these appear to cover about 98 per cent of requirements. The residual cases are fortunately easy to spot by the indexer at the pre-computer stage, and if missed they do not disrupt the other data-processing operations. They could probably be programmed but only at the cost of a considerable increase in the complexity of the present programme.

The adoption of the computer assistance in April led to a considerable rise in the productivity of the indexing staff. Some preliminary practice runs in using the special coding needed for the inversion references had been undertaken in 1967, and there were virtually no staff difficulties in coping with the change. After four months computer-assisted operation, the former backlog of work had been eliminated.

At the moment, for temporary technical reasons the extension of computer assistance to cover computer typesetting is in abeyance, but detailed plans for this further development are well advanced.

---

**INDEXERS' STATIONERY**

As a result of Mrs. Lantz's article in the last issue of *The Indexer*, at least one enquiry has been received as to the availability of such stationery as she described. Anyone, especially in the U.K., who is using stationery other than the normal 5 x 3 inch, or similar, cards or slips, is invited to inform the Hon. Editor of the material used and sources of supply. Such information will be made available to anyone enquiring.