Computer-assisted indexes: two results assessed

Hazel Bell and Kate Suggate

We hear much of the advantages to be derived from using computers to assist index-production. All of them are to the compiler, saving time and labour. There has as yet been less emphasis on the finished product of these processes; on any advantage to the user. As experienced assessors of indexes submitted with applications for admission to the Society of Indexers' Register, we applied a similar scrutiny to two indexes known to be, and indeed publicized in the press as, prepared with the help of two separate computer systems. (These were the only grounds upon which these two indexes were selected for examination.)

We have not succeeded in obtaining full technical details of these two systems, and cannot comment on the division of responsibility for the results; whether the programs were properly applied, or the programmer properly instructed by the original indexer. Nor do we know of any restrictions imposed by the publishers. We look merely at the final products presented to the world and vaunted in the press.

* * *

The first system examined is that of Alpha Data Services, used to prepare the index to Mea Allan’s delightful book, William Robinson 1838-1935: father of the English flower garden (Faber, 1982). Miss Allan wrote jubilantly of this to The Author (Summer 1982) under the heading, 'Hurray for the Computer Age!':

'I thought that fellow-biographers and historians might like to know of my experience with the piece of magic which has made the indexing of my forthcoming book simplicity itself. All the author need do is mark the page-proofs (I did mine with a red dot) to indicate the required entries (or you can leave selection to the indexer) [not further explained—Ed.]. Miss Computer does the rest, even unto putting it all into alphabetical order! Magic? Surely. No more juggling with index cards or slips of paper. At the touch of a button the indexer's worst headache is removed, and within a few days all is typed ready for the printer.

'I was a guinea-pig for the Alpha experiment. It worked.'

Alpha Data Services are advertised in their open letter, 'For authors, writers, and publishers', which includes the paragraph:

'Alpha Data Services have developed a method of INDEXING using a micro computer. This eliminates many weeks of tedious manual work—for example, we recently completed the indexing of a 250 page autobiography in less than eighteen working hours for the author’s first initial copy. This included a considerable amount of both technical and cross reference work. The index is prepared entirely to the author's individual specifications using the printer's final proofed manuscript.'

We asked for technical details. A letter from Alpha Data Services explains further, 'Our system is based upon the author's knowledge of the book and their judgement as to what should and should not be included in the index and what requires cross referencing.'

We examined with much interest this index, of 13 double-column pages for a 237-page text, guinea pig for a new computer system, the product of 18 hours' work subsequent to the marking of the text with red dots by the author. And we found that by the standards of admission to the Society of Indexers' Register, the Alpha experiment had not worked.

* * *

Alphabetization with speed is reputed to be one of the strongest points of computerization of indexing; but even here we find many abominable errors.

Cremation Society, The, 129-30
Cremation movement in Canada, 212
First crematorium in Canada, 212

Curtis's Botanical Magazine, 23

Curtis, Charles H., 223-4

George, Ernest, 178, 187-8
George, V, King, 219
George, Lady St, 19

Manchineel comes before Manchester; Virgin's Bower, The before Virginia; White Mountains before
White, Colonel; and under Robinson, subheading Periodicals, Flora and Sylva comes before Farm and Home, Woods and Forests before Wood Fires . . .

There is current divergence of opinion as to the correct way to treat M’, Mac and Me; but surely no system would advocate what appears here:

MacDonald, Ramsay
M’Ken, Mark John
Mackay, J. T.

Style is incorrect as well as inconsistent. There should be no punctuation, or a colon, after a main heading with no page references, not a comma as used here. Capitalization is inconsistent: compare

California, 100, 105, 152
flora of, 107
trees of, 107

with—

Garden, The, . . .
Colour in
Merges with Homes and Gardens
New editor appointed
Use of English names.

Subheadings are ill chosen, and often the significant word is not brought to the front. Here are some strangely listed ones from under ‘Gravetye Manor’:

Comes alive again
Gives up farming
Making the playground
Old mare and carter
Principles of landscaping
Redesigning fields
Interior

and from under ‘Robinson, William’:

Learns French
Love of children
Love of trees.

Entries are strangely hidden as well as dispersed. Robinson was much concerned with children. To find passages about these you must look under Gardening Illustrated, then the subheading, ‘Articles for children’; and under Robinson, subheading, ‘Love of children’ (5 references). Hardy flowers are found under Garden design as a sub-subheading under ‘Use of alpines’. The Bluebell Line is found only under Gravetye Manor, subheading ‘Planting railway embankments’. There is no heading, Railways, although the United Pacific Railroad is also featured in the text. (It appears in the index under America, subheading First transcontinental railroad.)

Under ‘Paris’, after an initial string of 15 page numbers, comes:

Parks, 74
After the Siege, 120-1

. . .
Parc Monceau, 61, 121
Parc des Buttes Chaumont, 63
Parks and Gardens, 61, 66, 103.

There are errors which can be categorized only as sheer confusion. For instance, the whole sequence

Turner Bailey, Messrs., 65
Turner, Richard, 37
Rudge & Turner, 234
Jack Mitchell, 234
William, 169

(‘Messrs.’ occurs in the text with ‘Turner Bailey’, but not with Turner, Rudge and Turner, auctioneers; so the computer preserves the ‘Messrs.’ in the one case, does not introduce it in the other. ‘Rudge & Turner’ seem the strange Christian names of the second Mr Turner listed. Jack Mitchell is not Jack Mitchell Turner, as might be supposed, but ‘Jack Mitchell of Turner, Rudge and Turner, East Grinstead’, the auctioneer. ‘William’ on page 169 is represented only by, ‘[Robinson] greatly admired painters like Crome, Corot and Turner’—in fact, Joseph Mallord William Turner.

Very many gardens are described in this book; but ‘Gardens’ in the index yields one reference only, to the Alpine garden at Verrières, on page 70, which reads in full, ‘Louis’s son Henry was an enthusiastic student of conifers, and of alpines—he was to found the famous Alpine Garden at Verrières’. ‘Verrières’ in the index leads only to page 68, and ‘Alpine gardening’ to ‘see Garden design’.

Books are cited both under author’s name and their titles, usually with different page numbers under the two entries, thus:

Garden Beautiful, The, 188 [entered between Garden Annual and Garden design]
Garden, Beautiful, The, 169 [commas sic. Entered between Garden tools and Garden, The ]
God’s Acre Beautiful, 129-30, 135
and under ‘Robinson, William’:
Garden Beautiful, The, 169, 182.
God’s Acre Beautiful (Cremation and Urn Burial), 129-30.

Alpine Flowers has 10 references plus one subheading under A; 5 references only under Robinson. The full title in fact should be Alpine Flowers for English Gardens.
The cross-references form a whole new field of error:

'France: Exposition Internationale (1867), see Paris'—under 'Paris' is listed, 'Exposition Universelle'—no 'Internationale'.

'Alpine gardening, see Garden design'. And look well. Alpine references are hidden there among 17 subheadings:

Alpine gardening, 55
Carpeting under trees, 95
With alpine plants, 157
Use of alpines, 82.

Cross-references for Alpine gardening are lacking to two subheadings under Gravetye Manor, Sussex:

Alpine meadow, 173, 179, 223, 227, 234
Walls and use of alpines, 148

as well as to

Alpine garden at Verrieres

as the only subhead under Gardens, and

Alpine plants, 157, 214
Alpines, 29-30, 33, 103.

Fruit culture, with 9 page references and one subheading, lacks either page or cross-references for Fruit culture as a subheading under both Gravetye Manor and Paris.

The introductory note states only, 'References to plates are in italics': but does not explain that the italicized numbers refer not to the page on which the illustrations appear, but to the plate numbers themselves, although the 37 photographs are all printed on pages of the text, not as a separate section or all together. The 10 line drawings have a separate numbering system, typographically ignored in the index and not always included in the correct page references. For example, Figure 1, page 74, 'The Napoleon Peach Tree', is indexed neither under N nor P, but under 'Fruit culture'. Peaches do not get even a cross-reference, although they are discussed in the text.

There are many strings of undifferentiated page numbers. Robinson's own entry begins with 18 of these, with another 19 under the subheading 'Principles of gardening'. There are 23 under Garden, The; 10 under Wild Garden, The (one of Robinson's books, not defined as such in index)—though admittedly only 4 under this book given as a subhead under Robinson. Similarly, his book, English Flower Garden, The, has 16 references when cited under its own (undefined) title, but a mere 7 as a subheading under Robinson.

Under 'Robinson' is a subheading 'Cremation'—leading not to details of his own cremation, but to his interest in cemeteries. Under 'Cemeteries' appears this strange amalgam:

'All the author need do is to mark the page-proofs to indicate the required entries . . . Miss Computer does the rest.' An index to a book which is more than technical listings, having true literary content and abstract ideas, can be compiled properly only by continual checking from the text to the index. This was not done for this index, and really, it shows. Perhaps this program is too primitive to cope with a book rather than a booklet, or the suggested method of marking the copy is incorrect; so that in the end the index to Father of the English flower garden proves a veritable bed of weeds. It includes many insignificant mentions-in-passing while omitting much interesting information. There is no discrimination between the relevant and the trivial; there are many inaccuracies; entries are placed where they would not be sought. It is inelegant, mechanistic, and difficult to use; altogether far below professional indexing standards; we think, below standards that should be accepted by a publisher.

* * *

The second computer system, Pastest Service, is developed by Farestead Associates of Hemel Hempstead, Herts, 'Indexers of Information'. In 'An open letter to indexers' they describe it thus:

An inexpensive computer programme is now available, which takes the routine work out of indexing. You may be able to save 30% of the total time it takes to prepare an index. The programme runs on an Apple micro-computer.
Alphabetical listings of headings, and sub-headings are produced automatically. You underline the items to be indexed in the normal way. When you are ready, you enter them via the keyboard with the page number. A simple code is used to indicate sub-headings. The computer sorts the entries and prints out the index. The programme enables you to correct mistakes in typing, and there are other error checks. For example, entries cannot be made without page numbers.

The programme allows indexing work to be done before page proofs are ready. An index can be prepared, using the page numbers of the manuscript. When page proofs are available, the computer receives a series of instructions which converts the old numbers to the new ones.

You can benefit from this new system in one of two ways. You can either buy the programme to run on your own machine, or one to which you have access, or you can use the secretarial service we are setting up.

The equipment you would need to run the programme, including the printer, would cost under £1500. The programme, plus a back-up service to sort out any software problems would cost under £80.

Alternatively, we are setting up a secretarial service to produce an index from a marked manuscript. A 24-hour turnaround will be guaranteed.

* * *

The Guardian, 1 July 1983, featured this business, reporting, 'They have already made use of the new system for indexing their latest book, Choosing and using professional advisers'—by Paul Chaplin, (Entreprise Books, 1983, £4.95). This was the second book whose index we examined as a publicized example of the finished product of an index by computer program.

The book, of 112 pages, sports a 'General Index' of only 79 lines (supplemented by Indexes of Useful Publications and Useful Organizations) three of these being:

- Fire Authority, 75
  - officers, 75
  - regulations, 75.

This inadequate length of index is printed in single column only to make two pages—a publishers' restriction, or computer's inability to produce two-column output? There is ample space for double-column, and some entries do indeed require expansion. For example:

- accountants: all aspects, 2, 6–20
- banks: all aspects, 2, 20–38
- solicitors: all aspects, 58–71

There is a string of eight page-references to 'Small Firms Service'.

The final chapter, 'Sources of information', remains almost unindexed.

Neither is there any entry for 'Job Centres', which has two paragraphs under this heading in the text; nor for 'Personnel Specialists' (five paragraphs), indexed only under 'recruitment and personnel', nor for 'News agencies'.

The computer system has omitted also to leave spaces between alphabetical sections.

Alphabetization seems to be letter-by-letter, yet we find:

- micro-computers
- micro-computer consultants.

Neither is the index free from simple pagination error: 'Microsystems Centres, 86' should read '96'.

This has in fact proved a depressing task, assessing the products of computer-indexing systems.

We supposed that computers could not replace the human faculty for judgement, précis, devising of le heading juste. But it is disappointing here to find that even on the mechanical aspects of accuracy, alphabetization, consistency, cross-references, organization of headings, synonyms, the use of these specially-devised computer systems has not produced work of a standard acceptable to the indexing profession.

It should be stressed that what we have here is the use of a kind of indexing bureau by authors, and that there is no evidence that the computers in question were used by qualified or even competent indexers. Apart from the alphabetization (the quality of which depends on the program) everything depends on the indexing ability of the person using the computer.

There is a term used by computer-workers: GIGO, signifying, Garbage In—Garbage Out, conveying that good results cannot come from a computer incorrectly or inefficiently programmed. We cannot say whether the faults here lie with the computer, the program, the programmer, or the author/indexer who has sought to take advantage of the slogan 'Let us take the hard work out of compiling your index' by marking his manuscript using the prescribed guidelines; the hardware, the software, or the liveware. But it does seem that those seeking the help of computer programs to prepare indexes should beware. Buying a computer and/or program cannot be enough, without someone to provide the basic instructions for each individual index who is fully knowledgeable of the techniques of indexing. Marking what takes one's fancy with a red dot does not invoke a 'piece of magic'. This rough magic we must abjure.