Computer-assisted indexes assessment

I read with interest the admirably clear and objective account of the two computer-assisted indexes reviewed by Hazel Bell and Kate Suggate in the October 1984 issue of *The Indexer* (14 (2), 95–8), and should like to offer the following comments.

Alphabetization for the computer is one of the most difficult tasks which it is asked to perform in the field of indexing. What the mind does readily, filing a before b, Arabic numerals before Roman letters, colon before semi-colon are, for the computer, complex tasks requiring sophisticated programming.

Instructing a computer how to deal with the many different filing elements within headings demands the skills of both the indexer and the programmer. For example, the rules for MARC cataloguing state that Sir Thomas More should be coded $aMore$eSir$hThomas $dSaint, the Se indicating an addition between the entry element and other parts of the name and which is not taken into account in filing. In the index reviewed, George, Ernest files before George V because the computer has not been instructed to file Arabic numerals before Roman letters. *Curtis's botanical magazine* files before Curtis, Charles H. because the computer was instructed to file apostrophes before commas.

The problems of alphabetization have been overcome (the British Library’s index to the British national bibliography is a good example), but the resulting programs were the results of many years of research and cannot, as far as I am aware, be run on a microcomputer, like the Apple, which lacks the required memory capacity.

I consider that the time when the indexer can expect a rapid sort resulting in immaculate alphabetization, cross-referencing and arrangement of subheadings without a good deal of text editing is still some way in the distance.

Olwen Terris
Assistant Information Officer,
British Universities Film & Video Council

Please allow me to comment on the article, ‘Computer-assisted indexes: two results assessed’, by H. Bell and K. Suggate in the October 1984 issue of *The Indexer*.

The disappointment expressed by Bell and Suggate stems from two separate problems—(1) the author’s marking up of the proofs, and (2) the competence of the computer program to prepare the final index.

With respect to the first problem, the computer program is not to blame if the author omitted important entries on the proofs.

The problem of alphabetization by the computer seems to be at the heart of the writers’ disapproval of the results. The explanation of the process is that many computers sort data (alphabetize) on a literal letter-by-letter basis, and do so by comparing the numerical codes for each letter in one entry with those in another. The codes for punctuation marks (commas, etc.) are lower than for letters, and so a comma is ‘alphabetized’ ahead of any letter; a space occurs ‘alphabetically’ before any punctuation mark.

Most index programs use a ‘built-in’ sort for alphabetizing. It is possible to program the computer to avoid most of the errors thus created, but the programming is complex, and the resulting program runs very slowly compared with a straightforward sort.

It also appears, from the results quoted in the article, that no proofreading was done on the index after the computer had finished.

I assist my wife in the running of our indexing programs, and we have certain ‘tricks’ which circumvent the limitations of computer sorting. For instance, all capitals are alphabetized ahead of all lower-case letters. When a subentry should start with a capital, we enter it in lower case. In this way the computer sorts the subentry correctly, and it takes little time, when editing, to change this back to a capital.

Editing a sorted index can be done in two ways: (1) on paper by marking up a printed version of the raw, sorted index; and (2) by using a word-processor on the computer. It may interest readers to know that the editing-by-hand and subsequent typing is no slower than the computer-edited version, and may even be faster.

Gary Hall
Leeds

Alas, no shekels

In carrying out research for *The press in Britain: an annotated bibliography*, which I am co-editing, I had occasion to refer in the British Library to the *Index to periodical literature of the world*, which was published (originally under the title *Annual index of periodicals and photographs*) by *The review of reviews* between 1892 and 1903.

These volumes are remarkable not only as detailed indexes to articles etc (in some 150 periodicals), but also in containing potted descriptions—illustrated by editors etc—of journals then published in Britain, Europe and
America; in fact a truly encyclopaedic international coverage. A classified tabulation is appended.

But that is not all: in the Index for the year 1891 (published 1892), the famous W. T. Stead, who had started The review of reviews with Newnes immediately after leaving The Pall Mall gazette, has this to say:

‘Indexing is one of the most useful things in the world, but like many other good things it does not bring in the shekels. Some day we shall have endowments for funding a College of Indexers for the English-speaking world . . .’

He pays tribute to his compiler, Miss E. Hetherington, and her team, and that lady herself contributes a five-page article in the following year’s Index (for 1892, published 1893) on ‘The indexing of periodicals’.

For anyone interested and holding a BL reader’s pass, the shelfmark is P.P.6483.c.

G. D. H. Linton
Sevenoaks, Kent

Reader-precise?

I read the article ‘Censorship in indexing’ (The Indexer 14 (2), 105-8) with both interest and enjoyment. How lucky we are that we do not have to face more brutal methods of censorship than the type described in Sheila Intner’s excellent paper.

Would it be possible for a Librarian of the non-indexing variety to offer Ms Intner some advice on terminology? She twice uses the term ‘user-unfriendly’ (pages 107 and 108). I have always found this term rather unfriendly to hear; why not use ‘user-hostile’? I have not seen this latter term on paper but ‘hostile’ seems so much stronger than ‘unfriendly’.

Tim Cullen
Librarian, Slough Laboratory
Ministry of Agriculture, Fisheries and Food

Oula Jones, indexer to Bernard Levin, comments on his remarks in ‘Authors’ attitudes to indexes’, The Indexer 14 (2), 85-6:

Indexing Bernard Levin is both terrible and delightful, in equal measure. Delightful because I know he’s as concerned about the minutiae as I am (do you enter Diana and Duff Cooper under their noble titles?); because he’s On My Side (though even he can’t prevent publishers repaginating after the index is finished); and because he observes the courtesies of life in scrupulously acknowledging my work every time and sending me a copy suitably inscribed: but terrible because he reads the index manuscript with his horrid penetrating gaze and picks up all my little slovenlinesses—initials where full names should be, titles of nobility avoided, over-abbreviated entries, as well as my (I hope) occasional appalling mistakes (I’ve just confused Vladimir Ashkenazy with Stefan Askenase and invented a poet called William Herrick, but then I’ve Been Ill . . .) The fact that despite all this he kindly refers to me as infallible I find terrible too.

However, not all authors share Bernard Levin’s enthusiasm for the index. Recently I heard that a friend of mine was writing a book on Northern Ireland, and wrote to offer my services, gratis, as indexer. As we discussed the style of the index, it became clear that he only wanted a name index, because he felt it was not scholarly enough to merit a full annotated index: however, as I read the book, I felt it was an important source work on Northern Ireland, and persuaded him to change his mind. I produced a ‘proper’ index, with no strings and lots of subheadings, of which I was fairly proud, for once . . . then alas! Instead of the book appearing in hardback, it was decided to bring it out in paperback only, thus necessitating re-pagination of the index which was done, alas and alack, by the author, who promptly lost his nerve and removed all my subheadings, leaving, in one case, a string of 101 page numbers for one of the main protagonists. He also omitted the entire section dealing with the subject of the book.

Worst of all, he gave me an Acknowledgement, identifying me by my maiden (Northern Irish) name, as well as my more anonymous Jones . . .

Oula Jones
London W3

Discontent of an author with index

Nov. 30 [1934] The Index to the book [Ego] arrived. Look at once to see if my editor’s name is spelled correctly and can’t find it! Look for Lord Camrose. Not there either! No Forbes-Robertson! Begin to examine Index seriously and turn it over to Jock [secretary] who reports that scores of important names have been omitted. And this is professional indexing! Yet the fellow, without meaning to be funny, includes such items as

- Cutler, Kate, and cremation
- Dates, Seven thousand tons of liquefying
- Wesley, Ghost of
- Whipsnade, Accident at

I shall decline to pay seven guineas for work which Jock has to do all over again.

— from Ego 2 by James Agate (Harrap, 1936).