Computer-supported Indexing

T. D. C. Kuch

Describes a method of using a computer to assist the indexer by typing onto a computer terminal while the index is being compiled; the draft index is then computer-sorted. The various ways in which the computer helps the human indexer are described. The major drawback to the method is cost.

In the October 1973 issue of The indexer, Jeff Maynard described his experiences using a computer in book indexing. Previous writers have described "computer-assisted indexing", "computer-aided indexing", or "machine-aided indexing" as consisting of either entirely automated processes, or else automated processes supplemented by human editors.

Clearly, Maynard has described the opposite approach, with the human being more definitely in command of the indexing process, and his approach deserves a different name.

"Computer-supported indexing", then, involves using the computer as an "indexer's helper", in the way apprentice smiths were smiths' helpers in former days, although it remains to be seen whether or not the indexer's helper can ever become a journeyman, given its inability to journey beyond the length of its electrical cord.

My own experience with computer-supported indexing was similar in many ways to Maynard's, but different in some significant particulars.

The index was typed on a computer terminal, while being written, using the WYLBUR interactive text-editing system, straight through the collection to be indexed, issues 1-38 of a technical journal. Each typed line corresponded to one card of the traditional index in progress.

At the completion of this process the draft index was computer-sorted, then lines with similar or identical entries were merged, and the whole was edited on the terminal. The finished index was then supplied to the publisher in the form of a data set stored on magnetic disc, the normal medium of WYLBUR data storage.

The "indexer's helper" supported the human indexer in these ways:

- **Association** of words, word-roots, etc., or association of items for a given issue after sorting by subject, as an aid to editing, consistency-checking of orthography, treatment, etc. As an example, the command "LIST 'system'" immediately provided all lines containing 'system', 'systems', 'systematic', etc.

- **String replacement**. What indexer has not changed his mind in the middle of the index as to the precise word he needs for a recurring concept? The computer was able, for instance,
to change every occurrence of "facilities" to "capabilities" with the command: "CHANGE 'facilities' TO 'capabilities' IN ALL". Provided the word to be changed hadn't been mistyped (e.g., "facilities") somewhere, the indexer is assured that all occurrences have been found.

Sorting of lines. Maynard is too sanguine about this; computers can't sort in the human style, not because of technical limitations, but because the human sorter brings to his job a great deal of background knowledge. For instance, we are obliged by standard practice to sort "London, Jack" ahead of "London, England". While it is possible to make the computer aware of famous writers and famous cities, all possibilities cannot be anticipated, such as the hypothetical "Baltimore, William Calvert, Seventh Baron" preceding "Baltimore, Destruction of". Desperate measures, such as writing "London (England)" instead of "London, England" cause as many problems as they solve.

Nonetheless, the computer was able to do a rough sort very quickly, relieving most of the indexer's sorting burden.

Shorthand expressions. These can be typed to save time, for long terms and phrases that occur frequently. In final editing, these can be lengthened by string-replacement. In the same way that "facilities" can be changed to "capabilities", "BSI" can be changed to "British Standards Institution".

Proof or draft listings can be produced in minutes. The devoted user of computer terminals, however, considers over-reliance on such "hard copy" to be inelegant.

Formatting was in this case up to the publisher, but a good text-editing system will allow one to vary line-widths, vertical spacing, depth of indentations, and to select the most efficient number of columns per printed page.

The final advantage of computer-supported indexing, not used in the present work, is that the entire index need not be sorted, formatted, or printed on paper at all, but merely made available to other users of computer terminals. "LIST 'microfilm'" would then extract and display every occurrence of that word, whether as a primary term or as a modifier.

The limitations of computer-supported indexing arise from the use of a general-purpose text-editing language not designed by, or for, indexers. Modifications, such as the ability to recognise or ignore differences of case or font at will, would improve the method considerably.

The major drawback of computer-supported indexing remains its cost. When the indexer is employed, however, and his time is valued at $10 to $15 an hour, a cost-advantage can generally be realised, given the substantial number of professional man-hours saved.

Notes:

(1) Jeff Maynard, "Indexing—A new approach", The indexer, 8, pp. 204-205 (October 1973).


(3) Two precursors of Maynard were Otto-Albrecht Neumuller, in an unpublished thesis (Göttingen, 1969), and Andrew E. Wessel, in "Some Thoughts on Machine Indexing", Santa Monica CA: RAND Corporation, Report No. P-3869, June 1968. Wessel argues for "machine-aided intellectual indexing, which can provide on-line error checking and error correction guidance, consistency checks, indexer training, and probably more thorough indexing by human indexers than is possible by purely manual methods". (Lawrence H. Berul, "Document Retrieval", in Cuadra, op. cit., 4, p. 218.


(5) Interface, 1968 to date, Bethesda MD: Division of Computer Research and Technology, National Institutes of Health. The completed index was published as issue number 39, dated December 30th, 1972.
