Microcomputer-aided production of indexes
Theodore C. Hines and Lois Winkel

Recent experience in working with microcomputers has led us to believe that they will play a major role in indexing in the future. They will serve both as production tools in themselves and as data collecting devices for indexes produced by larger computers.

Microcomputer equipment costs are already low enough to make purchase highly justifiable economically for the individual free-lance indexer who does any considerable volume of work. It already seems economically justifiable for publishers to acquire microcomputers for loan to free-lance indexers to use at home, as well as for data gathering or production use in the publisher's office. Equipment costs are likely to continue to drop and computing power to rise. A central processing unit microcomputer chip which cost well over $300 in 1974 now costs about $7.50 in quantities of one.

Background

This paper discusses some of our experiences with microcomputers which grew out of work with the establishment of a prototype Children's Media Data Bank at the University of North Carolina at Greensboro. We originally acquired our first microcomputer two years ago. It was purchased primarily as a terminal to be used with the university's paired IBM 370/165 computers and with a minicomputer, the Hewlett-Packard 2000 G. We bought it, rather than a regular terminal, because in stripped form (i.e. without much memory) it was not much more expensive than a terminal would have been.

It also offered the opportunity—as we could afford more memory—for introducing students to computing, for experimentation with microcomputers as input devices, and for generally finding out about the world of microcomputing.

It was more than a year before we were able to acquire sufficient memory to use the microcomputer as a stand-alone device. Since then we have been so impressed with its capabilities that we have a second microcomputer for the Children's Media Data Bank at the university, and will acquire more whenever funding can be obtained. We have also purchased our own microcomputer for home use. It is being used for the writing and editing of this paper, for example, and is also used for index and bibliographic production of a broad variety of types. It has completely replaced the noisy, slow teletype we formerly used as our home terminal. Since it is compatible with the university computers and microcomputers, it enables us to do programming, debugging and data collection at home.

The home computer

It is difficult to realize that the microcomputer in our home office is considerably more powerful than many second-generation computers long used for producing quite large information tools and services. It has the virtue of being completely dedicated to whatever use we are putting it to at any given moment. It has had practically no down-time to date. Its use is not subject to the quality and availability of telephone lines, nor does it involve telephone costs, except when we use it as a terminal. It readily gives us mylar ribbon impact printing quality as good as that from an IBM Selectric typewriter, with the same facility for changing type-faces which the Selectric provides—indeed, many terminal printers of this kind are Selectric based.

What we would like to describe here is the use of our microcomputer system for back-of-the-book indexes, together with a brief discussion of equipment configurations and costs. We will be quite specific about brand names and (approximate) prices, as an aid to other free-lance indexers who may want to consider microcomputer use.

Mainframe

At home we are at present using a Processor Technology Sol-20 microcomputer. We have about 65K bytes of immediately available memory (i.e. 65,000 characters of immediate memory). It is possible to make a good beginning with less memory, but we would not recommend less than 32K—and even this poses severe sorting problems. A good micro like the Sol with 32K of
memory will cost approximately $2,500. Additional memory runs about $150 for each 8K.

**Recorders and monitor**

Some money can be saved by using your existing cassette recorders and an existing TV monitor. You can even use (some) normal television sets by purchasing a video to radio frequency converter for about $35 or less—we have used the ATV Research Microverter with considerable success. Otherwise budget about $150 for a TV monitor and about $60 each for two cassette recorders. The cassette recorders must be of fairly good quality and have a tone control—we have had good results both with Radio Shack’s model CTR-21A and Panasonic recorders.

**Output equipment**

The last considerable expense is fairly rough—a good upper-and-lower-case output device. Good Selectric-based devices with which we have had some experience include those made by Andersen-Jacobson and the SelectTerm. The latter may be put under an IBM service contract for the typewriter part. Cost here can run anywhere from $850 (perhaps risky) to $2,000, but it might be possible to use one at a computer center while getting started. In that case it is interesting to note that it is easier to take the computer to the terminal than the terminal to the computer—our Sol is comparable in size and weight to an office typewriter. There are output devices which give proportional spacing and very high speed—like the Diablo, for example—but here we are talking in the $3,500 bracket and above and discussing output devices which may well cost more than the computer itself.

**Software**

Software is a considerable problem. We are lucky in that we are both programmers with considerable experience with information handling and systems design. In addition, for a number of years we have been developing unit operations for information handling similar to the unit operations concept in chemistry. These are essentially programming-language-independent operations common to most information manipulation problems. Some simple examples of such unit operations are paragraphing, paging and (perhaps not so simple) sorting.

This has allowed us essentially to escape any new systems design problems, and simply to translate our SNOBOL/SPITBOL programs written for the IBM 370 into BASIC and, in some cases, assembler language for the Sol.

Even so, we have found it advantageous to buy some software. A good text editor is almost essential. We have been most pleased with Michael Shrayer’s Electric Pencil Word Processor, which costs about $100. Its use in connection with our indexing package will be described below.

**Other hardware**

It would be possible to go into considerably more detail about equipment and software. We have at the university, for example, a Micropolis Metafloppy dual-disc drive which gives very rapid access to more than 600,000 characters. For this paper, however, we are concerned with illustrating what a minimal configuration can do as an aid to book indexing.

**The book indexing program set**

There are four basic parts to the book indexing program set. We use the Electric Pencil program mentioned above for initial data collection and editing. Input configurations are essentially very simple, but have been worked out quite carefully to minimize input effort and to provide entries in a form suitable for logical computer sorting.

**Input formats**

We keyboard from marked page proof. Page references are keyboarded only once for each page and automatically posted to each entry for that page unless the entry is a cross-reference or is already followed by inclusive pagination. Where a data element from one entry is repeated exactly from the previous entry, it is not rekeyed. Instead, the ditto sign (double quote) is substituted for it. A typical entry group might appear as follows:

```
43
Cats 43-51
Cats SEE ALSO Dogs
"—Veterinary medicine
Dogs—"
```

This example would be expanded by the second program in the set to:

```
Cats 43-51
Cats SEE ALSO Dogs
Cats—Veterinary medicine 43
Dogs—Veterinary medicine 43
```
Some other conventions which cut keyboarding include placing a plus sign before entries which, in our nomenclature, should be ‘flipped’. An example might be:

+ Cats—Veterinary medicine 43

which would become two entries:

Cats—Veterinary medicine 43
Veterinary medicine—Cats 43

The entry:

Cats—+ Veterinary medicine 43

becomes:

Cats—Veterinary medicine 43 and
Veterinary medicine 43

Similar conventions may be devised for particular cases, including Precis-type manipulations. The Electric Pencil program provides an automatic substitution feature which is quite useful. (This had also been part of the original SPITBOL programs.) If a word or phrase is likely to recur frequently, we simply key a unique abbreviation for it and the Pencil will expand all occurrences of that abbreviation to the full form. For example, in keying a recent list of journals we substituted 'j' for 'Journal of' at the beginning of a line and for 'Journal' in the middle of a line, and all examples of these abbreviations were expanded by a single instruction for each word. In keying a list of law journals alone, similar abbreviations saved us an estimated 30% of the keystrokes which otherwise would have been required.

Index vocabularies

It is not generally realized that the vocabulary of a particular index or bibliography is likely to be quite restricted, and that long words may appear with much higher frequency than in general running text. Thus abbreviation for later computer expansion may effect very great savings. We are indebted to James May (see bibliography) for putting our intuitive observations on this question on a sound basis and for much practical advice in this area. May’s concepts apply not only to keyboarding but to internal computer storage of data, a topic on which we are preparing a separate paper jointly with Dr May.

Entry development program

The functions of the second program of the book indexing group in posting page numbers and expanding entries, which we refer to as the entry development program, have already been indicated. This program is applied to the output of the Electric Pencil.

Sort/merge

The third program is a sort/merge which automatically produces appropriate sortkeys, and sorts or merges index entries by them. As mentioned above, sorting is definitely a non-trivial operation as far as index and bibliographic entries are concerned. The topic is discussed extensively elsewhere (see bibliography), but it should be noted that sorting takes considerable time on a microcomputer, as well as careful management of merges. We have found it best to do quite small sorts, merging at frequent intervals. This is especially true when working with cassettes.

Basic index program

The last program is perhaps the key program of the set. It consolidates page references for identical entries, handles the subordination and indention of subdivisions and modifiers, provides ‘continued’ statements when a subject is broken between columns or pages, creates pages and provides running heads and page numbers. It provides a manuscript index ready for the printer or suitable for direct offset (or other) reproduction.

A simple start

The Electric Pencil serves not only as our data-collection program but may be used to edit material on the video screen at any stage of production of the index. Our program set as described here is quite complex, and took a considerable time both in original development and in its translation from large-scale computers to the microcomputer. A great deal can be done, however, in book index production with only a sort/merge program and a good editor. Thus it is possible to start out with a quite simple system and gradually develop a more complex one like that described here as programming time is available and programming skills are developed.

Conclusions

While it may not be true of others who do not do so much writing, we have felt that the use of the editor program alone is probably sufficient to make ownership of a microcomputer highly justifiable economically—the thought of going back to an ordinary typewriter with its ac-
companying correction tape, opaquing fluid and cut-and-paste corrections is quite horrifying.

After our year's experience with microcomputers we have certainly concluded that working with them is most worthwhile and rewarding—as well as, on occasion, quite frustrating.

Bibliography
May, James H. Identification of design principles for efficient computerized bibliographic data input. (Doctoral dissertation, School of Library Service, Columbia University, 1978—to be published.)

Wheatley Medal, 1978

Only eight indexes were submitted to the joint panel of the Library Association and the Society of Indexers as candidates for the Wheatley Medal for indexing for 1978. Most served well the purpose for which they were compiled and none showed gross inaccuracies, but equally, the Panel, after considerable deliberation, decided that none of the submissions was of high enough standard or showed sufficient depth in its subject approach to merit the making of an award.

Of more concern to the Panel than the quality of the indexes submitted was the poor response from the two professions and from publishers in putting forward works for the award. Indexes, although they may not sell books as such, are often mentioned by reviewers and are also noted by cataloguers and used by hard-pressed reference librarians: yet few librarians, indexers or publishers make nominations for the award, which was instituted with the aim of encouraging the improvement of indexing standards. The Panel recommended that further publicity should be given to the Medal through the Library Association, the Society of Indexers, the National Book League and the publishing and bookselling world: it would welcome nominations (addressed to Michael Yelland at the LA) for an outstanding British index for 1979—one that is comprehensive and comprehensible, accurate, consistent, clearly laid out, suitable for the work to which it is a key, takes appropriate account of subjects and not merely names, and is usable by specialist and hurried consultant alike. Such indexes exist and have been given the award in the past and the panel hopes that the lack of such an index in 1978 will be remedied in 1979 by a larger field of selection.

R. A. CHRISTOPHERS,
Chairman, Wheatley Medal Joint Selection Panel.

Wheatley Medal, 1979

Entries are invited for the Wheatley Medal, 1979, from members of the Library Association and the Society of Indexers, publishers and others.

The Library Association Wheatley Medal is awarded annually for an outstanding index first published in the United Kingdom during the preceding three years.

Printed indexes to any type of publication may be submitted for consideration provided that the whole work including the index, or the index alone, has originated in the United Kingdom.

The final selection is made by a committee consisting of representatives of the LA Cataloguing and Indexing Group and the Society of Indexers with power to co-opt.

All recommendations should be submitted on a form available from the Library Association, to the Development Secretary, the Library Association, 7 Ridgmount Street, London WC1E 7AE, not later than 1st February 1980.

Congratulations

It is with much pleasure that we note that the first presentation of the H. W. Wilson Company Indexing Award has been made to Hans H. Wellisch for the index to his own book The conversion of scripts: its nature, history and utilization, published in 1978 by John Wiley and Sons, who also received a citation. The index was 'especially commended for its clarity, precision, and appropriateness in both its intellectual content and its typographic form'. Dr Wellisch is a member of the American Society of Indexers and a contributor to this journal.