Letters to the Editor

All your ideas, comments, accounts of problems and methods of solving them, concerning indexes and indexing, are welcome. Please send them in.

Computers: how much assistance?

Persistent confusion must be first resolved and then eliminated, or it wrecks its milieu.

Confusion persists in our Society's thinking about computers. We are confusing the warehousing of statuary with the practice of the arts of sculpture. (Note the plural, which strengthens my analogy.) So that Angels shall not fear to tread here, may The Fool step in to offer some observations, ask a question, and make a recommendation?

During the Saturday afternoon of our 1st International Conference at Roehampton four seminars were held, and the one at which I assisted (at least in the French sense) spent more than half its time talking about computers, although these were not named in the title given to the 26 of us for our discussions. Very nearly all of what was said I had heard several or many times before during the past seven or eight years at London meetings of the Society. One valuable distinction was made, namely between computer indexing and computer-assisted indexing, but this was not discussed in any detail. Time and again there was confusion between 'information storage and retrieval' (warehousing) and 'book indexing' (sculpture). Sometimes this confusion was open and obvious, sometimes it was submerged but almost equally obvious.

The fact that a book index is an example of information storage is utterly irrelevant, because it is a one-off job uniquely augmenting its book and, at present, wholly useless in relation to every other book. (I refer, of course, to a subject index.)

Computers are very good at information storage and retrieval, and some of the arts of indexing are called in for the design of both operations. But, just as the reader of a book with a 'good' index has no need of any knowledge or even awareness of the arts of indexing when he is using the index, so are those arts without concern to the ultimate user of information from a computer. In other words, the programmer of a computer is a special kind of indexer, but for instance any astronomer by following a few simple rules can extract from a computerized store of astronomical information those facts that he needs—though the data may have to be further 'processed' before he can 'read' them.

To resolve one of these confusions once and for all I ask my question. Does anyone with more than two years' full-time experience of computer-assisted indexing believe that a computer will ever be able to index an ordinary book on a new topic? By 'ordinary book' I mean not a bibliography or a list of titles, but a progressive treatment of a subject in chapters comprising paragraphs made up of sentences: for example, a biography, or a textbook of chemistry, or a history.

To be specific: let us suppose that we have a biography of the 1st Earl Baldwin, and let us suppose further that its subject index is generally agreed to be a very good one. This index and the book are 'fed into' a computer by an expert programmer who has read many biographies. Could the computer, without further programming except for normal computer-routine matters and the feeding in of a new book, then produce an equally good index to an equally comprehensive biography of the 1st Earl Attlee? Or likewise for an index to 'The Organic Chemistry of Phosphorus' based on a computer-stored index to 'The Organic Chemistry of Sulphur'? Or an index to 'The History of 13th-century England' based on a computerized index to 'The History of 14th-century England'?

If the answer to my question is a clear No! (my own, albeit uninformed, opinion), then every member of our Society should be told so in the clearest possible way in the pages of our journal. If the answer is a clear Yes!, then it should be backed up with some pretty substantial facts relating to the kind of book I have specified and these should be published.

Assuming that the answer will be neither of those clear monosyllables I come, finally, to my recommendation. It is that a general meeting of the Society should be held, with two appropriately experienced and mutually unknown computer programmers participating, to discuss the limitations of the assistance that computers can give now and may be expected in the future to give to the compilers of the subject indexes to 'ordinary' textbooks. This discussion should be reported in full detail in The Indexer, not merely the conclusions drawn from it. We really cannot, I believe, afford to waste any more time with peripheral talk about computers.

Yours always hopefully,

Neil Fisk.
From Professor P. A. Samet, 
President, The British Computer Society.

Mr Neil Fisk has shown me a copy of his letter 
concerning Computers and Indexing, and asked 
for my opinions about the technical feasibility of 
using a computer to prepare the index of a book. 
After a long discussion with him, I think the best 
immediate answer I can give is the following. 

As I see it, the basic problem is the con-
struction of a list of the significant items which 
are to appear in the index. It is trivial (even if 
technically difficult) to construct a concordance, 
with the obvious refinement of omitting words 
like ‘the’, ‘of’, ‘and’, etc. Such a concordance, 
however, is of little help to the user of an index. 
The selection of significant terms, which may be 
words or groups of words in the text or some 
collective term describing parts of the text (eg 
‘family’ in a biography), requires knowledge of 
the subject and judgement by the indexer. It is 
certainly possible for an indexer to provide such a 
list, but what happens with unfamiliar or new 
terms?

A perfectly feasible technique would be a form 
of co-operation between a human indexer and the 
electronic slave, to give computer-assisted in-
dxing. The first stage of analysis would 
essentially be the construction of a concordance 
(with simple words removed). If a list of words 
that must be included is available, occurrences of 
these in the text could be suitably marked, for 
transfer into the index, but this is not an essential 
feature, merely a time-saver. The next phase 
would be the display for the indexer of words in 
order of decreasing frequency, for decision as to 
inclusion or rejection. In case of doubt it would 
be possible for the indexer to request the context 
to be displayed. In this way the list of index terms 
could be built up. A refinement would be for the 
program to add the results of the indexer’s 
decisions to its lists of terms to be accepted and 
rejected, for future use.

Many modern computer systems allow inter-
action by the user, normally from a console like a 
typewriter with a printing mechanism and often 
with a (much faster) TV-display. Such a system 
would make the operation I have described very 
r rapid indeed. A more expensive system would use 
a ‘light pen’ and a TV-display, with the indexer 
marking up the text as it is shown to him, rather 
like the traditional method.

The purely mechanical parts, such as sorting 
into alphabetical order, getting page references 
correct, and so on, are relatively simple computer 
operations. Perhaps the biggest problem of all is 
getting the original text in ‘machine readable form’; it may be necessary to type the complete 
text into a computer, it may be possible to use an 
optical reader, but this is expensive, or, best of 
all, a machine readable copy is produced as a by-
product of the original typing or from the type-
setting.

I hope your readers find these comments to be 
of some help. I am sure that several of my 
colleagues in the British Computer Society would 
be pleased to discuss the matter further, if 
required.

Yours sincerely,

Paul Samet.

Book indexing

I wonder whether I am alone in deploring what 
seems to be the increasingly frequent use of the 
expression ‘back of the book indexing’.

The words ‘back of the’ surely add nothing 
useful to ‘book indexing’? Literally understood, 
they would be absurd: it is not the back of the 
book that is indexed, nor does the indexer work 
at the back of the book.

Where these superfluous words may have some 
real effect is through damaging association with 
the many pejorative meanings of ‘back’—back 
stairs, take a back seat, backward, back row of 
the chorus.

I hope writers on indexing will turn their backs 
on this fashionable usage, and get back to plain 
‘book indexing’.

Yours truly,

Anthony Raven.

A publisher’s attitude to indexes

I do not know who gave you the information 
that we do without an index wherever possible, 
but this is certainly not our attitude. Although 
indexes are costly and large ones can be very time 
consuming, the costs of an index are insignificant 
compared to the other costs involved with a non-
fiction book.

Most of our non-fiction books published this 
year include indexes, and please be reassured of 
the high value in which we hold them and the 
services you indexers make available to us 
publishers.

Your sincerely,

Jeremy Greenwood.

Director, Trade Division, 
Cassell Ltd, Publishers.