

The impact of technology on indexing

Jan Ross

The indexing process has changed remarkably with technological advances. Indexing is no longer just 'back-of-book' indexing, but includes automatic indexing, machine-aided indexing, web indexing and even 3-D indexing. Not all the effects have been positive, especially for the indexer, but the future of the Internet and efficient information retrieval lies with indexing.

There is no doubt that the impact of technology on indexing has been considerable, if not revolutionary. From the use of cards and shoe-boxes we have progressed, depending upon your viewpoint, to PC-based indexing programs, automated indexing, machine-aided indexing and web-indexing. Whole new paradigms for indexing are being developed. On the other hand, the technology has had potentially negative effects – indexing is frequently deemed to be superfluous or costly in the era of free-text searching.

Dedicated indexing programs

To elaborate a little – the development of dedicated indexing programmes such as MACREX and CINDEXT in the 1980s was a major step forward for most professional indexers.¹ No longer was it necessary to spend hours sorting and re-sorting (especially if the cat used the table as a launching platform) thousands of cards, and typing and re-typing the final index. The time-consuming and laborious chore of alphabetization gave way to increased time for the intelligent work of indexing. Indexes could be re-ordered from word-to-word to letter-by-letter, or even into page number order, at the press of a key. The locators could be re-numbered to cope with the inevitable repagination of the text, and indexes could be re-styled with commas, lower-case or bold. Indexing moved from an unsophisticated cottage business to a profession using sophisticated indexing tools. Indexing programs moved onto Windows-based platforms and were updated with new functions, and new programmes still continue to be produced, such as Sky Index and Idxtools. Doubtless there will continue to be improvements, such as the incorporation of better spell-checkers, thesauri and functionality for electronic indexing, though this is already being developed. CINDEXT has add-on programs for one-pass indexing of electronic text and hard copy (HTML/Prep), whilst MACREX now has integral embedding functionality for the embedding of index entries within SGML/HTML (and later on XML) text.

Automated indexing and machine-aided indexing

With the advent of CD-ROMs and the increasing popularity of the Internet, there have been new demands for indexing tools. The sheer volume of data on the Internet has necessitated an element of automation. The need for speed and

cost-reduction has forced the development of totally automatic indexing packages such as Indexicon. Whilst professional indexers do not rate such programs highly, nor the indexes so prepared, there is still a need for them, though this is limited to simple data structures. Automatic indexing is little better than full-text searching, and has the same inherent problems: the relevance of the term is not considered. However, there is now software that does consider relevance. Semantic processing software, such as CoBrain, is 'intelligent' software which semantically processes electronic text files by reading them and analysing the terms, their relationships and contexts to produce a knowledge base, akin to an index. Perhaps this is a way forward – yet another impact technology has had on indexing – but as a professional indexer, I do not believe that the semantic processing role of the computer can be used in anything other than very structured and subject-limited texts.

To improve upon the inadequacies and weaknesses of fully automatic indexing, machine-aided indexing programs have been launched, such as Machine-aided Indexer (MAI). Here, speed and efficiency are brought to the process of indexing using a natural processing tool, or 'knowledge basis' (a progressively built thesaurus, to you and me), but the human brain is still required to maintain accuracy and flexibility. We have all had the experience of being deluged with 'hits' during a free-text search and yet still missing the important results. A machine-aided indexing program avoids the pitfalls of automatic indexing by scanning the text and providing a list of suggested subject terms matched against an internal thesaurus to exclude inconsistencies and prevent editorial drift, so that the (human) indexer is able to fine-tune the index terms and index structure. Machine-aided indexing ensures consistent indexing over a long period of time, but combines this with the benefits of speed and cost savings of automatic indexing. In much the same way as dedicated indexing programs have taken the mundane sorting tasks away from the indexer, so perhaps this will be the next step forward, allowing computer assistance for the selection and wording of index entries. It is certainly being used for some very large databases, such as the US National Library of Medicine's MEDLINE.

Website and 3-D indexing

Both indexing across the Web, and indexing of individual websites, are other areas in which technology has had a huge

impact on indexing, though space prevents a discussion of the former and of search engines. There is, however, a need to index individual websites, and this need will increase as websites increase in size and complexity and as more and more data are published on them. This has resulted in specific web indexing software being developed, such as HTML Indexer. Since the development of the first web indexing packages a few years ago, new programs have now been marketed including a standard-based automatic indexing program for the indexing of meta-data or databases (Z'mbol). Such programs use structured text, so are unsuitable for free text. Recently, a new program has been launched that allows not only the indexing of websites but also invisible web sources; that is, the information held within non-HTML databases (including directories and dictionaries) that are available to websites (InvisibleWeb). The pace of development of the Internet is quite frightening, but it is becoming clear that although more and more data are becoming available via the Internet, the need for accurate and efficient indexing is becoming ever more important. Technology is, ironically, not only raising the profile of indexes and providing a new context for them, but is also raising the profile of indexers. Indexes, and indexers, are becoming more essential for efficient information retrieval than ever before.

Not only do websites need indexing, there is also a trend to publish large encyclopedias and other free-text titles on the Internet. The concept and design of indexes has to become broader to make full use of the new technology and functionality of the software. Users of electronic indexes now expect 'add-on value'. One of the ways that this can be provided is by developing concept mapping, or as I prefer to call it, 3-D indexing. These are indexes that ostensibly have many of the same properties as standard 'back-of-book' indexes, but have a three-dimensional element in that the cross-references are all hyperlinked in such a way as to create a web of interacting entries. The user can migrate up and down the index, across (i.e. sub-entries and their equivalent), and also backwards and forwards (or even diagonally) via cross-references. Indexes are becoming complex webs. As a portal is the entry route to many related websites, so the index is becoming an intrasite-portal.

New working methods

Even indexes for printed books and journals have not been excluded from the effects of the technology revolution. Indexes are now provided on disk in formats suitable for immediate typesetting, and can even be sent by email across the world in minutes. No longer will the excuse 'it's in the post' be adequate! Email has transformed the delivery of indexes and communication with clients. Indexing is also becoming more international with the advent of the Internet and email, and so our market is growing. So, too, are the demands. The 'instant society' now wants instant indexes, with pressures on the indexer that are becoming excessive.

Technology has not only affected indexing and indexes, but also indexers at a personal level. Indexers are becoming skilled computer operators, in some cases even programmers. They are learning to apply technology in ways

unthought of ten years ago. On the downside, we are becoming wedded to our computers, with all the resultant problems of repetitive strain injury and back problems and, I believe, we are losing the ability to communicate at a human level. How often do we send an email instead of talking to a colleague or client? Then, of course, there is the stress factor that I alluded to above – definitely an adverse effect of technology.

From the use of a cardboard box 20 years ago, the impact that technology has made on indexing has been nothing short of amazing. The indexing process has been transformed, and the transformation process is likely to accelerate with the increasing size and importance of the Internet. Indexing is changing – even the very meaning of the word is changing. 'Indexing' is becoming blurred with 'information retrieval', though admittedly that is the primary function of indexing. Even search engines are described as 'indexing' the Web – but anything less like an indexing process, in the true sense of the word, is hard to imagine. Indexing, by professional indexers, now makes use of ever-more complex and advanced programs, with the development of more and more innovative 'indexes'. The Internet is pushing information retrieval processes forward, but by building on the past. The future is indexing, but not as we know it now!

Note

- 1 The indexing programs mentioned are examples only; no attempt has been made to provide a complete listing. No endorsement of the programs is intended, nor are full details provided. Please contact the respective companies for further details. Websites are listed below.

Software cited and sources

CINDEX™ – www.indexres.com
CoBrain™ – www.cobrain.com
HTML Prep – www.LevTechInc.com
Indexicon™ – www.iconovex.com
Idxtools™ – www.sid.cam.ac.uk
InvisibleWeb – www.invisibleweb.com
Machine Aided Indexer (MAI)™ – www.dataharmony.com
MACREX™ – www.macrex.com
SKY Index™ – www.sky-software.com
Z'mbol™ – www.fdgroupp.com

Jan Ross, a postgraduate in the medical sciences, has been indexing full time for over 21 years in the fields of medicine and food science. She was awarded the Wheatley Medal for an outstanding index in 1993 for the indexing of Encyclopedia of food science, food technology and nutrition (Academic Press), and again in 1997 for Rheumatology (Mosby International) – a unique double! She is a director of Merrall-Ross International Ltd, a company providing publishing services to medical publishers, indexing training, medical writing and distance-learning material development, and electronic publishing consultancy. Jan is also involved in developing indexes and innovative information retrieval methods for directories, encyclopedias and other major publications to be published on the Internet. She is an active member of the Society of Indexers. Email: m.r.i@btinternet.com