Indexing—ways of working, now and into the future

Glenda Browne

Professional indexers create indexes of many sorts for many different clients, including publishing houses, government departments and big businesses. Although there is one general approach to indexing there are many variations in the way we work depending on our own styles, our clients, the work and the budget.

A general approach

On receiving a brief, indexers familiarise themselves with the document to get a feel for the scope of the project, and to detect any potential problems. Familiarisation includes some or all of:

- checking that every page is present

- reading the table of contents and some sample sections of the work. At this stage indexers who feel that they cannot index the work might return it to the client

- reading the whole book. Few indexers have the time to do this, although Collison (1972: 26), who lived in more leisurely times, says 'the book should first be read through two or three times as a whole'!

- marking up the page ranges for all sections, for example, noting on the proofs that a section runs from pages 15 to 17.

Indexers then go through the page proofs highlighting important words and phrases, and making notes of synonyms and other features of the terminology to be used. To decide whether a term or concept should be highlighted they ask themselves 'If I looked up that term in the index, would I be pleased to find this information?'

The next step is to enter the chosen terms, following the style that was provided or chosen. Most book indexers use specialised indexing packages such as SKY Index, MACREX or Cindex—these automate much of the filing and format for the index.

Editing the index to ensure consistency of headings, to organise subheadings and to add and check cross-references can easily take thirty per cent of total indexing time.

The index is then printed in word processing format and read through again, ideally by the indexer and another person. When final changes have been made, it is usually emailed to the client.

'Ride a wild pony' vs. 'dressage'

Sonya Hartnett (2004), an author who now uses a structured method for developing a plot, says she used to use the 'ride a wild pony approach' in

which she would throw all the bits of the story together when they came to mind. This nicely describes my way of indexing. As a matter of some urgency, I like to get all my provisional content into my indexing software. Then I can relax and take my time to optimise the wording and structure. However, as I plod though my revision tasks I sometimes think it would have been better if I had thought a bit more about the structure as I went, so I did not have to spend so much time editing at the end. Hartnett calls this her 'dressage' stage, when everything is tightly controlled.

Sherry Smith (Smith and Kells 2005) describes how she delays decision making about complex areas of the index until her final editing stage, by which time things have often resolved automatically. She calls this process enlightened procrastination (which sounds a lot more professional than 'ride a wild pony'!).

To mark up, or not to mark up

Some indexers do not mark up the text, but enter terms directly into indexing software. They may or may not have read the book first, although they are almost certain to have examined the table of contents and other introductory material. Some indexers only mark page ranges (for chapters and sections) in advance. In effect, they perform the analysis, term selection and entry steps at the same time. The advantage of this approach is that it can be a lot quicker; the disadvantages are that concepts may be missed, and the significance of content early in the book may not be fully understood until the end of the book. For straightforward, highly-structured books the method works well.

Number of passes

Indexers also differ in the number of passes they make through a book. Occasionally indexers with abundant time available read the book, then highlight important concepts, then enter the terms, then edit the index. This is a luxury most indexers working to a deadline cannot afford.

Most indexers highlight terms in one pass through the book, then enter terms in another pass. Others enter terms without marking up, as described above. Other indexers make a number of passes through the book.

For structured computer books, Robert Saigh (2004) indexes chapter heads followed by text heads and any subheads on his first pass. He then starts again and indexes any italicised, hyphenated or boldfaced terms. If he has any space left, he selects terms from the text. He says: 'My process mimics Ford's automation in that I do discrete passes, which do not overlap.' If space is limited it can be useful to index level by level as he describes. Linda Fetters (1996) describes a similar approach for indexing what she calls the forest and tree entries. She recommends writing the general (forest) entries for a chapter first to build the index structure, and then to add in the specific (tree) entries.

Sherry Smith (Smith and Kells 2005) also makes at least two passes through the book. She often indexes scholarly books, and finds that the first pass allows her to determine the main structure of the index, while the second pass lets her pick up points she has missed the first time, and allows her to refine her wording and structure. The first pass replaces the highlighting step, and offers the advantage that some structure is being developed so she can start examining new terms in context.

Large team indexing projects can also benefit from a number of passes, allowing some issues to be resolved before the whole index is completed.

Working with PDFs

Working with PDFs and walking on water are easy if both are frozen.

Source unknown.

PDF is a universal file format that preserves the fonts, images, graphics and layout of any source document, regardless of the application and platform used to create it. PDF documents can be viewed and printed using free Adobe Reader software.

Publishers now often email PDFs to indexers instead of posting hard copy. Indexers may:

- work from the electronic document (sometimes on a different monitor to the indexing software),
- print hard copy, and absorb the cost, or
- print hard copy, and charge the publisher (about 10c per page).

Instead of emailing PDFs, some publishers provide the indexer with access to a website or FTP (file transfer protocol) site at which they, and all others working on the project, can access the PDF files. This adds time to the work of the indexer who has to seek out the documents. More importantly, these documents may be tinkered with after they have been loaded to the site, so the files the indexer accesses one day may be different the next. Ideally this should not happen: if it does, the editor has to inform the indexer, and the indexer may have to charge an additional fee for changes that have to be made.

Using PDFs can also cause headaches with pagination. Some documents are initially sent unpaginated (for example, for an encyclopedia), and in others the pagination displayed by the software and the pagination written in the book do not always agree because the software includes pages which may be unnumbered, or numbered with Roman numerals. Indexers who are sent multiple files often combine them into one, while others convert them into different formats (for example, using File > Save As Text, or a progam such as Able2Extract).

Access to PDFs can also be beneficial as they are searchable, and can help you find topics you missed. But this can lead to overindexing if indexers select every occurrence of the topic without discrimination. PDFs can also be used for 'cut-and-paste' of terms into indexing software.

Repagination and updating

Sometimes indexers have to change the page numbers in an index when the page numbers of the document have changed. This can happen to a few pages during the initial indexing of a book, for example, when a table has been moved within a chapter. It can also happen to entire books when the format changes, for example, from hardcover to softcover. Most indexers detest the work, but some who specialise in it and have developed efficient systems find it acceptable, and say it usually pays well. Editors who expect to be publishing in different formats should consider embedded indexing to avoid the need for repagination.

To repaginate, indexers:

*   get the old book, the new book and the old index in page number order (either using an electronic copy, or scanning it in)

*   work systematically through the old index adding the new page numbers. If chunks of pages have moved exactly to new pages, they can simply increment the page numbers. Indexing software can do this automatically, but you should start at the bigger numbers or your new page numbers will be the same as some old page numbers

If it is not a simple movement of whole pages to new page numbers, indexers will have to mark the new page breaks on the old pages and change the page numbers one at a time.

Updating an index involves the addition or removal of index entries as well as repagination of existing entries. There are four main types:

*   Fixing mangled jobs (almost impossible!).

*   Yearbooks. These retain most of their structure from year to year, and are often adapted throughout the year by editorial staff.

*   Revised editions. Whether to redo an index or adapt an old one depends on the extent of changes, how easy they are to identify and whether the old index is available electronically. It is usually quicker to reuse an original index if the additions, deletions, and rewrites total less than twenty per cent of the book, especially if the indexer is the same person who indexed the previous edition.

*   Unfinished work. This may be needed if the original indexer is unable to complete a job. It involves completing an index in the same style as it was started. If not much work has been done it may be better to start afresh. If you start afresh you can still use some of the work that was done on the existing index by importing it into software such as SKY Index and then deleting it. This leaves the existing entries available as prompts for typeahead, but does not impose them on the new indexer.

Non-English indexing

Indexing practices are different in different countries, and for different languages. Multilingualism is entrenched in constitutions such as those of

the European Union, Canada and South Africa, so attention to multilingual index access is important. South Africa has a small publishing industry, publishing mainly in English, and also in Afrikaans and the nine main Black languages.

French publishers interviewed by Bella Weinberg (2000) believed that a minority of non-fiction books need indexes and that authors are the best people to create them. It appears as if most of the indexes that are published in France are generated as simple keyword indexes using word processor concordance software. Indexes are expected by many French-speaking Canadians—the fact that many books published in France do not have indexes does not mean that French books in general do not.

In Germany and other European countries, book index layout (for example, of subheadings) and filing order of certain groups of letters (Ll in Spanish, and Sch and St in German) are different from those in English-speaking countries.

Non-Roman scripts need special filing rules. For example, Chinese indexes are filed either according to the number of strokes in the radical element of the characters, or according to the order the words would be in if transliterated into Roman characters using the Pinyin system (Walker 2000). Some African languages such as those of the Khoi-San (Bushmen) and Xhosa have sounds such as clicks requiring the use of characters that may not be readily available in software, nor easily searchable.

When books are translated into other languages, the indexes can be machine translated or professionally translated, or a new index can be created. Machine translation is sometimes used for technical books, but does not work as well for books with less precise content or clear structure. Professional translation works better, but depends on the translator understanding the indexing terms out of context. Sharon Wright (now Sharon Bower, Index-L, 10 January 2005) gave the example of the indexing of the word evacuation in the phrase hurricane evacuation routes as bowel movement in Spanish!

Translations often take more or less space than the original documents—if text has flowed to different pages or has been cut to fit the space available, the perils of repagination are added to the problems of translation. The ideal is to have the work re-indexed from scratch.

Future indexing—e-books

For books available electronically, text searching is replacing indexing in some cases.

A usability study by Barnum (2004) found that users got more accurate results using the electronic back-of-the-book index, but preferred using text search to find information in PDF documents. This is probably an instance of the paradox of the active user, which says that people have a 'production bias' and like to jump into activity with a system, rather than preparing carefully so that they get the most out of it.

Text searching of e-books is improved by the use of passage level indexing, a type of partitioning applied to full text documents, which reduces spurious relationships and improves retrieval (Lancaster 2003: 190). Williams (1998) reports improved precision with a moderate decline in recall when full text searches used 500-word overlapping window passages from the Computer Science Technical Reports collection of the New Zealand Digital Library (NZDL). Another project found improved access through e-book indexes that reorganised conceptually—that is, when the search was done, a subset of the index that was considered relevant to the search was presented to the user on one page, allowing them to browse all headings and subheadings from the index that had fitted the search criteria (Chi et al. 2004).

Some mobile phones now include e-book functions. While this may increase the use of e-books, it is hard to picture indexes playing a major role in information access on such small screens (it is also difficult to picture effective full text search).

Indexing has traditionally been an important part of publishing. It is currently being challenged in some areas by the electronic provision of information, while in other areas it has become more important because people are struggling to find the content they need. There are sure to be many changes in the next decades, and indexers will continue to work towards providing tools that help readers find the information they need.

References

Barnum, Carol, Earvin Henderson, Al Hood and Rodney Jordan, 'Index versus full-text search: a usability study of user preference and performance' , *Technical Communication* 51:2, 185–206, 2004.

Chi, E. H., L. Hong, J. Heiser and S. K. Card, 'eBooks with indexes that reorganize conceptually', in *Proceedings of the Human Factors in Computing Systems Conference* (CHI2004) conference companion, Vienna, Austria, 2004, www.clsp.jhu.edu/research/malach/pubs/ JCDL2002MALACH.pdf.

Collison, Robert L. *Indexes and Indexing*. 4th rev. edn, London: Ernest Benn, 1972.

Fetters, Linda K., *Handbook of Indexing Techniques: A guide for beginning indexers*. Rev. edn. Port Aransas, Tx.:FimCo Books, 1996. There is also a 3rd edn. published by Fetters Infomanagement Co. in 2001.

Hartnett, Sonya, 'The colour of success' *Australian Author* 36:3, 2004.

Lancaster, F. Wilfred, *Indexing and Abstracting in Theory and Practice*, London: Facet Publishing, 2003.

Saigh, Robert, 'Index or perish' , *Key words,* 12:4, 123, 129, 2004.

Smith, Sherry L. and Kari Kells, *Inside Indexing: the Decision-making Process*. Bend, Oregon: Northwest Indexing Press, 2005.

Walker, Alan 2000. 'Impressions of the Dalian Conference of the China Society of Indexers' , *The Indexer* 22:1, 21–22.

Weinberg, Bella Hass, 'Book indexes in France: medieval specimens and modern practices' , *The indexer*, 22:1, 2–13, 2000.

Williams, Michael, 'An evaluation of passage-level indexing strategies for a technical report archive', *LIBRES: Library and Information Science Research* 8:1, 1998, libres.curtin.edu.au/libre8n1/williams.htm.

THE FINE PRINT