

Software, hardware and the web: reviews

Edited by Maureen MacGlashan and Nancy Mulvany

Indexing for technical communicators, Max McMaster and Sue Woolley. Vermont South, Victoria, Australia: Monarch Computing Services, Nov 2004. ISBN: 1-921055-00-6: \$35.00 (online book, www.monarchcomputing.com.au)

This review is based on the CD-ROM version, which contains the 140-page book as a PDF file, together with four Exercises and their answers as either PDFs or MS Word documents. The package is also downloadable from the Monarch website.

Max McMaster is a hugely experienced professional indexer, but writing an indexing guide for technical communicators is a challenge for anyone. It requires a delicate balance between patronizing authors with too little, and boring them with too much, detail, while avoiding the opprobrium of fellow professionals for over-simplifying. Perhaps we're even slightly uncomfortable, as a profession, with all attempts to instruct authors in our craft? But Mr McMaster's fellow author, Sue Woolley, contributes her expertise in IT support and technical communications, and it's good to be able to report not only that together they have avoided all three pitfalls, but that I was frequently delighted by the ease with which they reduced apparently complex topics to their essentials. The book's layout is clear and attractive, the development is logical, and examples are ample and well chosen. I had problems resolving the finest detail in a few of the screenshots on my PC using Acrobat 6.0 Professional, but this had no significant effect on overall comprehensibility.

The authors begin by stressing that indexes are a tool for the reader, explaining their access advantages over tables of contents and full-text searching, and urging the special importance of indexes to online help files. They then stress the importance both of planning indexes before embarking on them, and of editing them once they have been generated, before progressing to an overview of MS Word indexing (as the commonest and most widely available embedding system). There is a briefer section on using online help authoring packages. Their targets indeed seem chiefly to be manuals and online help text – both areas notoriously badly indexed in the past – and Mr McMaster well understands the need to index problems and tasks, not the relevant operations and options, in the latter. The book is full of sound common sense and excellent advice, sprinkled with some forthright assertion about indexing rights and wrongs, one or two of which might just raise the odd eyebrow, while the authors' estimated times for indexing operations certainly made me swallow hard, but the former at least are exactly what their intended audience will be looking for. And yes, readers are urged in certain cases to seek outside professional help, rather than index their own work.

Although concepts like set-out and run-on styles and embedded indexing receive passing mentions before being fully explained, this is presumably why the authors have provided a glossary (it includes all three). I found it slightly more confusing that the text instructions for inserting *see* and *see also* cross-references when embedding in MS Word are identical, and that the treatment of page spans doesn't actually invoke the pre-defined bookmark, although in both cases the screenshots provided ought to guide readers to the correct solution.

After the glossary, the book ends – of course – with an index, which is surprisingly light (at only just over two pages it even violates the authors' own rule of thumb suggesting '4%–5% of the length of the book') but seemed highly effective whenever I consulted it. If this

book, with its no-nonsense approach and practical advice on indexing, were required reading for all technical communicators, the world would become very slightly a better place.

Bill Johncocks, freelance indexer and member of SI

TExtract. Book indexing software. Harry Bego, Utrecht :TEXYZ. Single-user price ranges from €75 for a single-publication licence for the standard version to €1950 for a permanent licence for the professional version. A free trial download is available at <http://www.texyz.com/texttract/download/index.htm>

I am strongly in favour of anything that simplifies or speeds up indexing, and I was pleased to be asked to review TExtract, which creates indexes from pdf files in a format suitable for further editing. My expectations of automatic subject indexing are very low, and I would never use an automatic indexing program to do more than list words which might be useful for index-checking, locating missed proper names or specific terms, or as starting points for an index of proper names.

The website offers a full description of the program, plus some glowing tributes from users, mainly indexers of legal publications in the developing countries. Here is what it says it does:

Automatic creation of the initial index, with automatic construction of index terms:

- compound index terms ('pre-existing legal relationship')
- prepositional phrases ('security of transactions')
- automatic rotations ('promissory estoppel, doctrine of ~')

In-context selection and editing functions using parallel display of index and text:

- topic-oriented index views
- quick navigation and inspection of references
- global and local selection of entries and instances
- in-place editing and adding of entries and instances
- supports both index-oriented and text-oriented index making methods

Flexible input and output formats:

- input documents in PDF or paged plain text
- single and multiple-file input documents
- customisable output format in rtf
- easy compliance with publishing standards.

TExtract is certainly easy to use: you just fire it up, tell it what document to use, select name, subject or full index, and then if you want to, select various other options, or just let it get on with the job.

I made a pdf file of the final draft of ISO999, first removing the long section on form of names, and inserting a table of names and towns of six indexers. I used ISO999 (the international standard for indexing) because many readers will be familiar with it, and its index, and even those who have not seen it will have some idea of what it might contain. I chose to make a name index first, and then a full index.

TExtract works fast, and presents a draft index in tabular form, with statistical information and the facility to examine each entry in context by clicking on the locator, similar to the Acrobat search

facility. In this format you select entries and edit them, and then TExtract will make a final index in rtf format.

This is all excellent. One does not expect automatic indexing software to produce a usable index, or even a very intelligently compiled one, but even an alphabetical list of all capitalized words can be very useful for checking whether one has included all references to a particular person, or all the systematic names in a botany index. A list of strings of capitalized words would be even more useful. Of course one can use Acrobat search, which allows one to see found words in context, but it is a very tedious activity.

So, anticipating a very long list of capitalized words and phrases from an automatically generated name index, I opened the edit screen. Below is what I got, converted to TExtract's final index format, with my comments added against each entry. I have not removed a single entry, but the program only allows export up to the letter H.

INDEX

[no entries under A]

B

Bits 16 – [Bits & bytes (newsletter)][not included in full index]
Bonaparte 17 – [Bonaparte, Napoléon see Napoléon I, empereur des Français][not included in full index]
Burkina Faso 19 – [got that right in both indexes]

C

Chinese 28 – [ISO 7098:1982 Documentation – Romanization of Chinese][not included in full index]
Collison 16 – [R L Collison, in example][not included in full index]
Congress 28 – [Library of Congress][not included in full index]
Consistency 2, 8 – concept
Context 3 – concept [not included in full index]
Coverage 10 – concept [not included in full index]

D

Discriminate 6 – concept [not included in full index]

E

Elliptic 14 – concept
England 21 – [Bank of England]
Euripides 26, 27 [but not Aeschylus, Antigone or Aristophanes]

F

Fashions Club 22 [hard return tab]Club 147
Fashions{hard return tab}Club One Holidays] [not included in full index]

G

George 21 [Bagby, George]
George Bagby 21 [Bagby, George] [not included in full index]
Geschichte 11, 12 – concept [Geschichte (historische Entwicklung) Geschichte (Romanliteratur)]
Geschichten 12 [OK if it wasn't a concept] [not included in full index]

H

Hazel Bell Hatfield 7 [from a table of 6 names and towns] [not included in full index]

NOTE: This is a trial project - the index is exported up to letter H.

[Created with TExtract © Texyz 2005]

I rather expected it to find most of its entries made from the ISO999 examples, rather than the narrative text, but surprisingly it

did get the incorrectly included concept entries from the narrative. Oddly enough, it did include Antigone and Aristophanes in the full index, but not Aeschylus.

The full index was quite a bit longer, and contained rather more names than the name index. In fact it looked quite professional, apart from the capitalized concept headings, until one began to look closely. As promised, it contained phrases:

Alphanumeric arrangement
Assimilating non-verbal headings
Entitled Index 1

and some inversions:

Arrangement 1-3, 6, 8, 21, 22
hierarchical ~ 14
ordered ~ 5
Case letters
lower ~ 12, 21
upper ~ 12, 21

It never found Christine Shuttleworth or Drusilla Calvert, but this time it did find Ian Crane, and indexed him under I. I had hoped it would also find the word 'crane', but oddly from an example in French and English it picked out 'grue' (the French word for crane).

In short, you cannot assume this software is going to pick out words or phrases on any predictable basis, and while it is easy enough to delete what is superfluous and edit what is incorrectly structured, you cannot know what it has omitted.

If I were beta-testing this software, I could probably work out why it was missing some things and including others, and even missing words in one pass and finding them in another. However, as a reviewer, all I can do is warn against the use of unreliable tools.

Janet Shuter

WordEmbed. External index embedding tool. James A. Lamb. www.wordembed.jalamb.com. £64.62, €94 or \$110 (online payment via Paypal)

Around three years ago, discussion within the SI Future Group concluded that the best aid to embedding index entries in source documents would be a tool that inserted a unique number in the document, and linked these numbers to headings in a professionally compiled index. At that time I wrote a document for BSI IST/41, which has responsibility for XML development, and BSI IDT/2 which has responsibility for index development, outlining possible ways to proceed. The real purpose of the discussion, and of that document, was to put down a marker for indexers, and to make the case that instead of trying to invent ways to replace indexers, programmers would be better occupied designing tools to promote their efficiency.

It is a real pleasure now to review a tool that does precisely what we want, does exactly what it says it will do, and does it well. The website says:

WordEmbed allows the professional indexer to index a Word document in their own, familiar software and then, with a press of a button, turn the result into an embedded Word index. Creating embedded indexes now need take hardly more time than a conventional index.

Advantages are:

Creation of Word bookmarks for text ranges and their inclusion in XE tags is fully automatic
Handles tables correctly – copes with Word using the same page & line number for two places in one document

Software, hardware and web reviews

No need to drag each heading into the document manually
Work on an uncluttered screen (no Word formatting characters shown)

See clearly on screen to where locators refer

Define ranges both by creating a range locator or by using two point locators

Normal, Bold, Italic and Bold/Italic locators taken across automatically

Headings/cross-ref texts are all taken across fully formatted

Supports multiple sequences (multiple indexes) for one document e.g. Author Index, Subject Index

In practice, WordEmbed functions as a Word toolbar, not as a separate program, and takes only a couple of minutes to install and even less time to learn to use.

You need to have the Word source document open, and beside it your compiled index, open in Wordpad or Notepad or whatever, and in either MACREX-style string format, or hierarchical CMS-style format. All you have to do is highlight the point or range to which you want to attach a Word XE index field, and use a key-combination to bring WordEmbed into use. WordEmbed highlights the selected area, creates an XE field with a temporary location tag, and places the tag in the Word clipboard so that you can immediately paste it against your index heading in the index file. The tag is no more than a unique number identifier in plain text.

When you have finished doing that (remembering to Save both files frequently), you can select another WordEmbed function to insert all the headings from the index file into the correct XE fields in the source document, and compile the Word index. There really is no more to it than that.

However, in the real world, that is the point at which you find you need to do a little more index editing. Lamb suggests you work in Word to make changes to the XE fields which now contain your index terms and from which the client will generate the final index. Of course that will work, but I should be inclined to make a copy of the temporarily tagged source document file, run the WordEmbed embedding routine on that, make any necessary changes to the original index file, and run the WordEmbed routine again on another copy of the tagged source document. This is much simpler to do than to describe, and much preferable to searching through the source text for the XE fields you want to change.

The whole beauty of WordEmbed is that you can do what you like to the headings in your index file, so long as you don't lose the unique number WordEmbed tags, and so long as you maintain either the string format or the CMS format.

In fact, since I disapprove of visiting a source document more than once, I should not be inclined to go through the source document making the index in the usual way and then, having finished the index, go back to the source over again and try to work out where on the page I wanted to embed each index term. I normally do a good deal of pasting from the source document to my index file anyway, in order to save typing, and so I should do the whole heading construction/temporary XE operation in one pass. In fact I tried this, and it works, and this way you really can make an embedded index as fast as you can make a conventional one.

Finally, the website says you need Word 2003 and Windows XP, but I worked successfully on my old laptop in Word 98/Windows 98SE.

Janet Shuter

FreePint newsletter. 1997– Subscription free at
<http://www.freepint.com/subs/>

FreePint is an online community of over 72,000 information workers from all over the world. Membership is free, and details of

the range of services provided can be found at <http://www.freepint.com>. In this review, I want to focus on just one of those services: the twice-monthly FreePint newsletter.

Perhaps the first thing to say about the newsletter is that it rarely, if ever, contains anything specifically about indexing. It does from time to time deal with topics of interest to freelance businesses: there have been feature articles on time management, marketing, website hosting and website usability, for example. Though this is insufficient to justify describing the publication as essential reading, I do recommend it as one worth skimming regularly, especially by those who like to keep an eye on developments and concerns across the broad domain of information retrieval and provision, of which indexing is part.

Skimming is easily done, as the contents list at the beginning of each issue carries enough detail to enable quick identification of items worth closer inspection. Non-members of FreePint may read the newsletter (including archived issues) online, but one of the benefits of joining the community is that you can arrange to have each issue delivered by email, as soon as it is published. Or, if you prefer, the service will send merely an email notification that a new issue is available online.

The content follows a standard pattern, beginning with an editorial by FreePint's founder, William Hann. Sometimes this will inform readers of new or improved services available from the FreePint website; sometimes it will discuss a current issue of relevance to information provision and retrieval generally. Next is 'My favourite tipples': five favourite online information sources, with links and brief explanations of why the author (different in each issue) finds them particularly useful.

'FreePint bar' derives from the online forum of that name. This section highlights discussions at the bar since the appearance of the previous issue, particularly questions which, at the time of publication, have not yet been answered. 'Info – Jobs in information' derives from another FreePint service, listing some of the posts advertised and linking to further information about each, as well as to the full list of jobs.

A 'Tips article' offers advice and pointers to resources on some aspect of professional information work. The articles cover diverse topics: examples from 2005's issues include professional networking online, website usability, health information resources and job hunting. The book review which follows will also reflect the breadth of the FreePint community's interests. Among titles featured in recent issues are *The web library: building a world-class personal library with free web resources*; *Online information services in the social sciences*, *Find it online: the complete guide to online research*, *Information auditing: a guide for information managers*, and *Practical copyright for information professionals*.

Then comes the newsletter's main feature article. As with a Tips article, it may be up to 1,500 words long, but is even more diverse in coverage. Blogging, business awards, freedom of information, recycling, child rights information, competitive intelligence and putting archives on the web are the subjects of recent articles. Elsewhere in the archives you will find those I mentioned at the beginning of this review, plus ones on scenario planning, the semantic web, copyright and much else. I would recommend the one on RSS feeds (issue 161), solely for its wonderful opening line, but I suspect *The Indexer's* editor would rather I directed you to issue 166: 'Writing for the profession'.

Finally, the newsletter looks to the future, and to the past. 'Forthcoming events' lists conferences, seminars and meetings in the UK and elsewhere, while 'FreePint gold' takes a 'look back at what FreePint covered at this time in previous years'.

Linda Sutherland, freelance indexer and librarian, and member of SI