

Software tools for indexing: revisited

Nancy C. Mulvany

Basic features of four embedded indexing programs are discussed and compared with a dedicated indexing program. The programs are Word, FrameMaker, PageMaker, RoboHELP, and Macrex. The tests performed duplicate tests done ten years ago. The author outlines features that have improved during the past decade and deficiencies that still exist in embedded indexing software.

Ten years ago I spoke at ICEBOL4, the Fourth International Conference on Symbolic and Logical Computing, about programming issues and indexing software. An article based on that presentation appeared in the October 1990 issue of *The Indexer* (Mulvany 1990). In 1989 the indexing functions of four popular word processing products were compared to a dedicated indexing software product. These programs were Microsoft Word, WordPerfect, WordStar, XyWrite, and Macrex. The indexing modules in the word processing programs were found to be deficient in several respects. The article concluded, "Given that the index is a very important part of any book-length document, it is indeed unfortunate that current text processing software has little to offer in the realm of embedded indexing." Suggestions were offered for improving the sorting routines, index formatting, and user interface design.

What has happened during the last ten years? Quite a bit has happened. In the personal computing industry ten years represents several generations of CPUs, operating systems, and software applications. We have witnessed increased development of the Internet and the expansive growth of the World Wide Web. In the realm of text processing great strides have been made in the development of sophisticated features in word processing and page design software. In the realm of indexing, the University of Chicago Press published a new edition of *The Chicago manual of style* (1993). Some of the standards referred to in the 1990 article no longer exist. In fact, the only standard on the presentation of indexes is ISO 999 (1996). Dedicated indexing software products have added more features and have become more powerful. Today indexers are asked to work with text printed on paper, text in electronic files for printing on paper, and text presented online.

With the help of Peg Mauer and Jan Wright, four embedded indexing products were evaluated. These products are Adobe FrameMaker 5.5, Adobe PageMaker 6.5, Blue Sky Software RoboHELP WinHelp 7.0, and Microsoft Word 97. I compiled the results for Macrex 7.04 (see the References for software publishers' Web site addresses). FrameMaker and PageMaker are page design products; they can also produce documents in an online, hypermedia format. Blue Sky's RoboHELP is used to produce online help files; we tested their WinHelp version. Word 97 is a word processing program. Macrex is dedicated indexing software.

Test Procedures

As in 1989, the primary objective was to evaluate the ability of these programs to produce an index in the proper format with a minimum of ease on the part of the indexer. The indexing chapter in the 14th edition of *The Chicago manual of style* is without a doubt the most widely used style guide for indexes in the United States. The very basic style requirements of the University of Chicago Press that we evaluated are:

- letter-by-letter alphabetization
- leading function words in subheadings are not sorted
- bad breaks within entries should be fixed

In addition, some indexer usability features were also evaluated. These include features such as the ability to copy an entry, auto-invert heading levels, create entries while viewing the indexing in sorted order, and

the ability to directly edit the indexing with automatic updating of entry tags. The findings about these features are summarized in the “Index Entry and Editing” section of Table 1.

Each evaluator was given the same list of 26 sample index entries that were used in 1989. They were asked to type the entries in without any modification. In other words, if the program had the ability to force a sort order, they were instructed not to use it if extra keystrokes or mouse movements were required. The sample entries are all quite simple. Indexers using any of the dedicated indexing software programs would type these entries directly without modifying them.

Table 1: Features of the programs analysed

Features	Word	FrameMaker	PageMaker	RoboHELP	Macrex
Alphabetizing					
Word-by-Word	Yes	Yes	Yes	Yes	Yes
Letter-by-Letter	No	Yes	No	No	Yes
Reference Locators					
Allows for multiple locators per entry (dogs, 34, 89)	Yes	Yes	Yes	Yes	Yes
Combines page references (1, 2, 3, becomes 1-3)	No	Yes	No	N/A	Yes
Cross-references					
Allows for “See”	Yes	Yes	Yes	Yes	Yes
Allows for “See also”	Yes	Yes	Yes	Yes	Yes
Verifies cross-references	No	No	No	No	Yes
Hypermedia indexes only: cross-references automatically linked to target entries	N/A	Yes	No	No	No
Sorting of Entries					
Hide a character from the sort	Yes	Yes	Yes	No	Yes
Force a sort order	Yes	Yes	Yes	No	Yes
Ignores leading function words in subheadings	No	No	No	No	Yes
Index Entry and Editing					
Copy an entry	Yes	Yes	Yes	Yes	Yes
Copy a portion of an entry	Yes	Yes	Yes	Yes	Yes
Auto-invert heading levels	No	No	Yes	No	Yes
Ability to create entries while viewing the index in sorted order	No	No	No	Yes	Yes
Ability to directly edit the index with automatic updating of entry tags	No	No	Yes*	Yes	N/A
Index Output Features					
Produce indented style	Yes	Yes	Yes	Yes	Yes
Produce run-in style	Yes	No	Yes	No	Yes
Fix bad breaks (widow/orphan control; “continued lines”)	No	No	No	**	Yes

*Automatic updating of embedded tags only works in the Show Index box; it does not work if the compiled index is edited.

**Please see the discussion about Index Output Features in the text.

Table 2: How the programs alphabetized the sample entries

Word 97	PageMaker	RoboHELP	FrameMaker	Macrex
"Of Mice and Men"	data base	"Of Mice and Men"	"Of Mice and Men"	data base
data base	data entry	data base	data base	database
data entry	data structure	data entry	database	data entry
data structure	database	data structure	data entry	data structure
database	features	database	data structure	features
features	and protocols	features	features	of host mode
and protocols	memory usage and	and protocols	and protocols	memory usage and
memory usage	of host mode	memory usage and	memory usage and	and protocols
of host mode	form	of host mode	of host mode	form
form	form of input	form	form	format
form of input	format	form of input	format	FORMAT statement
format	FORMAT statement	format	FORMAT statement	form of input
FORMAT statement	Intel 80386	FORMAT statement	form of input	Intel 8088
Intel 80386	Intel 8088	Intel 80386	Intel 80386	Intel 80386
Intel 8088	Müller Bridge	Intel 8088	Intel 8088	Müller Bridge
Müller Bridge	Mzarak Turnpike	Müller Bridge	Müller Bridge	Mzarak Turnpike
Mzarak Turnpike	"Of Mice and Men"	Mzarak Turnpike	Mzarak Turnpike	OF MICE AND MEN
Of Mice and Men	OF MICE AND MEN	Of Mice and Men	OF MICE AND MEN	"Of Mice and Men"
OF MICE AND MEN	Of Mice and Men	on line	Of Mice and Men	Of Mice and Men
on line	on line	onliners	Of Mice and Men	onliners
onliners	on-line	onion skin paper	onliners	onliners
onion skin paper	onliners	online	onion skin paper	onion skin paper
online	onion skin paper	on-line	on line	on line
on-line	online	sleep therapy	on-line	online
sleep therapy	sleep therapy	Sleeping Disease	online	on-line
Sleeping Disease	Sleeping Disease		Sleeping Disease	Sleeping Disease
			sleep therapy	sleep therapy

Note: Shaded columns are letter-by-letter, the remaining columns are word-by-word alphabetizing.

Results

Most of the discussion in this section will be about the performance of the four embedded indexing programs. For years dedicated indexing software products, like Macrex, have provided the functionality needed by indexers to do their work properly. Please see Table 1 for a summary of features.

Alphabetizing

Table 2 presents the way the programs alphabetize index entries. FrameMaker is the only embedded indexing program that allows for letter-by-letter alphabetizing. FrameMaker also can sort the index entries in word-by-word order. It would be expected that the Macrex and FrameMaker letter-by-letter arrangements would be exactly the same. However, the sample entries were not alphabetized in the same manner. FrameMaker sorted the "Of Mice and Men" entry as though it began with a symbol and placed it above the alphabetic entries. Double quote and single quote marks are commonly ignored by dedicated indexing software. The other differences between the FrameMaker and Macrex results will be discussed below.

Word, PageMaker and RoboHELP all appear to be using the word-by-word alphabetizing method. Again, one would expect the results to be the same, but they are not the same. Word and RoboHELP, like FrameMaker, appear to treat the double quote mark as a symbol and sorted "Of Mice and Men" before the alphabetic entries. PageMaker handles the entry correctly and places it in the O's. RoboHELP, since it uses the Microsoft WinHelp compiler, stumbled over the entry, *OF MICE AND MEN*; it simply dropped it. The WinHelp compiler apparently considers *Of Mice and Men* and *OF MICE AND MEN* to be the same.

It is common in word-by-word alphabetizing to treat the dash or hyphen in the same way as a space is treated for sorting purposes. PageMaker handles this correctly, however both Word and RoboHELP placed *on-line* adjacent to *online* rather sorting it next to *on line*.

It should be noted that Macrex is capable of following all standard (and some non-standard) forms of alphabetizing rules. This includes conformance to University of Chicago Press and ISO 999 rules for both letter-by-letter and word-by-word alphabetizing.

Numbers within Entries

All embedded indexing programs failed to arrange the *Intel* entries in ascending numeric order. Unfortunately this repeats the same mistake found ten years ago. Numeric order is prescribed when numbers occur within the headings (Chicago 17.102; ISO 999 8.3.c)

Reference Locators

All programs are capable of merging entries that are the same but appear on different pages. FrameMaker is the only embedded indexing program that can combine page references into inclusive numbers. FrameMaker can change *books, 1, 2, 3* to *books, 1-3*.

Cross-references

All programs allow for “See” and “See also” cross-references. None of the embedded indexing programs provide sophisticated cross-reference checking that we have come to expect in dedicated indexing software. So, it is quite possible to create blind and/or circular cross-references with no warning from the programs.

Macrex, FrameMaker, PageMaker, and RoboHELP produce indexes for online (electronic) display. However, FrameMaker is the only one that links the cross-reference to its target entry. In a hypermedia index, it is a rather basic expectation that cross-references would be linked to the entries to which they refer.

Sorting of Entries

Word, FrameMaker, PageMaker, and Macrex can all hide a character from the sort and force a particular sort order. These are necessary features in indexing software. It is very common for indexers to hide characters so that they do not sort. Also, forcing a sort order comes up frequently as well. For example, the indexer may want to arrange the entry, *5th Avenue*, as though it was spelled out as *fifth avenue*. RoboHELP does not offer these features. It should also be added that these features are undocumented in Microsoft Word.

It is very common practice for leading prepositions, articles, and conjunctions (i.e., function words) in subheadings to be ignored for sorting purposes. The sort then operates on the important keyword following the function word. Because this practice is so standard, dedicated indexing software provides users with a built-in list of common function words to ignore when they are the first word in a subheading. Aside from Macrex, none of the programs tested provide this feature. They all sorted the leading function words in subheadings.

Index Entry and Editing

All the programs allow for copying an entry or a portion of an entry. Within the embedded indexing group, only PageMaker provides an easy function for the inversion of heading levels. That is, PageMaker allows easy swapping of main heading and subheading so that they reverse places and become a new entry.

As any indexer knows, it is extremely helpful to be able to create new index entries while viewing the existing indexing in alphabetic order. Working this way often reduces the amount of time spent editing later. If one entry is made for *files* and later another is made for *file*, this will result in two separate index entries. However, if the indexer can immediately see the previous *files* entry, then use of the singular term can be avoided.

Aside from Macrex, RoboHELP is the only program that allows the indexer to create entries while viewing the index in sorted order. Word, FrameMaker, and PageMaker do not provide this important feature.

All four embedded indexing programs insert some type of code in the text file being indexed. The ability to directly edit the formatted index with automatic updating of these codes is a huge time saving feature. RoboHELP and PageMaker provide this ability. Word and FrameMaker require users to locate the entry tags in the text files and manually change them.

Index Output Features

All programs can produce indexes in the indented format. Macrex, Word, and PageMaker can also produce indexes in run-in style. None of the embedded indexing programs can fix bad breaks within index entries. This deficiency is particularly troublesome for FrameMaker and PageMaker, two programs that are often used to produce book-length documents. Managing widow and orphan lines and inserting continued lines where appropriate is a basic necessity when designing an index for publication. Macrex has provided this feature since the early 1980s when it ran under the CP/M operating system.

Some may argue that no bad breaks exist in the online help indexes such as those produced by RoboHELP. While it is true that online help indexes are not presented on pages, it is still possible for the main heading to scroll off the screen producing a situation where a subheading is the top line on the screen. I consider this a bad break. Anytime it is possible to lose the frame of reference within an index entry, there is a bad break. If a subheading starts a column on a printed page or in a window on a computer screen, the frame of reference (i.e., the main heading) is missing.

A Decade Later

Has embedded indexing software improved in the last ten years? Yes it has. We now have a program that can alphabetize an index in letter-by-letter or word-by-word order (FrameMaker). One of the programs can follow the typical method of word-by-word alphabetizing (PageMaker). We now have a program that allows the indexer to add entries while viewing the index in sorted order (RoboHELP). Unlike the results in 1989, all the embedded indexing programs allow the indexer to copy previous entries. Three of the four programs allow the indexer to force a sort order or hide a character from the sort (Word, FrameMaker, and PageMaker); in 1989 none of the embedded indexing programs allowed for this. If we could blend the best features from all four programs we would be on our way to designing a competent embedded indexing program.

Unfortunately there are some very basic features still lacking in embedded indexing software. During and after the writing of an index cross-reference verification should be performed. This is particularly important after the index has been edited because cross-reference target entries can change. Several of the embedded indexing programs allow the indexer to hide characters from the sort, however indexers should not need to take extra data entry steps to hide leading function words in subheadings from the sort. The use of these words is so common that the software should handle them properly in an automatic manner. In deference to those who prefer to sort on leading function words, this feature should be user definable.

None of the embedded indexing programs can sort numbers within entries in ascending numeric order. This was true of the programs tested ten years ago. Since so many users of embedded indexing software work with technical documentation I find this blunder most curious. Technical documents often contain index entries that include numbers. It is amazing that this problem still exists.

As I stated earlier, the test procedures used had very simplistic requirements. Initially I wanted to add more complex requirements. However, it readily became clear that during the past ten years embedded indexing software has not even begun to keep pace with the feature enhancements in dedicated indexing software.

For years I have been using features like thorough cross-reference verification, the ability to group together index entries with the same characteristics, quickly check on the density of indexing, look at my index in page number order, or do global search and replace operations on the index. I take features like this for granted. These and other features combine to help improve index quality. The hard part of indexing is not using the software tools, the hard part is thinking and deciding about the content of the index. Software tools should help and, by all means, the tools should not get in the way.

Users of embedded indexing software may never be able to produce an index as quickly as users of dedicated indexing software simply because of the physical act of embedding the index entries. However, the publishers of such software could certainly add features that ease index entry and editing tasks. Indexers should not have to work in the dark, they should be able to see their index in alphabetic order whenever desired. Improvements in the alphabetizing and sorting methods should be addressed immediately. Numeric order has been part of our culture for a very long time. Perhaps as we step into the next millennium programmers of embedded indexing software will have learned how to deal with index entries that include numbers.

Although embedded indexing programs have improved since 1989, when compared with dedicated indexing software they are still crude tools. It is difficult to understand why so many basic problems still exist. In regard to alphabetizing and index format, *The Chicago manual of style* is readily available, it is not an obscure text. During the past ten years no less than three books about indexing have appeared, two of them have gone into second editions (Fetters 1994, 1999; Mulvany 1994; Wellisch 1991, 1996). All of these books discuss index structure. Given the availability of information about indexes and indexing it would be difficult for designers of embedded indexing software to claim ignorance. It appears that they are simply inept.

Acknowledgements

The author gratefully acknowledges the help of Peg Mauer and Jan C. Wright. Both are indexers and users of embedded and dedicated indexing software. Their willingness to perform various tests with embedded indexing software is appreciated.

References

Adobe FrameMaker. <http://www.adobe.com>

Adobe PageMaker. <http://www.adobe.com>

Blue Sky Software RoboHELP. <http://www.blue-sky.com>

Chicago manual of style (1993). 14th edition. Chicago: The University of Chicago Press

Fetters, Linda K. (1999). *Handbook of indexing techniques*, 2nd ed. Corpus Christi, TX: FimCo Books.

ISO 999 (1996), *Information and documentation --- Guidelines for the content, organization and presentation of indexes*. Geneva: International Organization for Standardization.

Macrex. <http://www.macrex.com>

Microsoft Word. <http://www.microsoft.com>

Mulvany, Nancy C. (1990). Software tools for indexing: what we need. *The Indexer* 17 (2) October 1990, 108-113

---- (1994). *Indexing books*. Chicago: The University of Chicago Press.

Wellisch, Hans H. (1996). *Indexing from A to Z*, 2nd ed. New York: H. W. Wilson.

Nancy Mulvany is the author of *Indexing books* and the owner of Bayside Indexing Service in Kensington, California. She teaches indexing, consults on the design of indexing software, and is a past president of the American Society of Indexers. Mulvany can be reached at nmulvany@bayside-indexing.com.