

Indexing software

Caroline Diepeveen, Middelburg (The Netherlands), Jochen Fassbender, Bremen, Michael Robertson, Augsburg

This article looks at the different types of software available for indexing and whether these can be used to produce the type of index that is required by international standards. As it is often claimed that indexing can be done by a computer and no human effort is needed, the authors first look into why indexing has to be a human activity and why automatic indexing cannot produce a satisfactory index. Automatic indexing will at best provide a starting point for further indexing. The authors then look into the possibility of indexing by using general-purpose software. This can produce acceptable indexes, although it may be fairly cumbersome. For the professional indexer, dedicated indexing software is therefore the preferred option. The three most widely used dedicated indexing programs are briefly reviewed. Their orientation and appearances are very different, but the features and final output are often very similar. Website indexing is a relatively new speciality within indexing. Two software programs used for website indexing, one freeware and one commercial, are discussed. The overall conclusion is that indexing is an activity that needs human intellectual input; software greatly enhances this input, but cannot replace the human input.

Indexierungssoftware

Dieser Beitrag handelt von unterschiedlichen Arten verfügbarer Software zur Erzeugung von Registern und untersucht, ob diese dazu benutzt werden können, ein Register nach internationalen Normen zu erstellen. Da oft behauptet wird, dass die Registererstellung mit einem Computer und ohne Einsatz des Menschen durchführbar sei, untersuchen die Autoren, weshalb Indexieren eine Aktivität des Menschen sein muss und weshalb eine automatische Registererstellung kein zufriedenstellendes Register hervorbringen kann. Automatische Registererstellung kann bestenfalls einen Ausgangspunkt zur weiteren Indexierung liefern. Anschließend wird die Möglichkeit der Registererstellung mit allgemein verfügbarer Software untersucht. Dies kann akzeptable Register hervorbringen, wenngleich oft nur auf mühsame Weise. Für den professionellen Indexierer stellt daher spezielle Indexing Software die bevorzugte Option dar. Die drei am meisten benutzten speziellen Indexierungsprogramme werden kurz bewertet. Ausrichtung und Aussehen dieser Programme sind sehr unterschiedlich, aber die Merkmale und Output-Optionen sind sehr ähnlich. Website Indexing ist ein relativ neues Spezialgebiet innerhalb der Disziplin des Indexierens. Zwei Programme – eine Freeware und ein kommerzielles – zur Erstellung von Registern von Websites werden erörtert. Das Fazit insgesamt ist, dass das Registermachen eine Aktivität ist, die intellektuellen Input des Menschen benötigt. Software kann den Input hervorragend verbessern, aber nicht den Anteil des Menschen daran ersetzen.

'There is nothing automatic about the index-writing process.'
Nancy Mulvany – *Indexing Books* (1994, p. 245)

Introduction

One of the most common questions encountered when telling people that one is a professional indexer is: 'But can't a computer do that?' The answer is that indexers do indeed use computers, but that, as with translating, the human intellect is very much needed to produce a sensible index. Selection of search terms can't be done by a computer, because context is everything. Professional indexers, therefore, use software to assist them in the indexing process. This article discusses the software tools that can be used for indexing.

After we have shown how 'automatic' indexing works, and why it won't produce a useful index, we discuss three different kinds of software tools that can be used for indexing. First, we discuss the use of software tools that most people already possess, such as word processors, spreadsheet and database software. These packages will typically be used by non-professional indexers and people who only have to produce an index occasionally, such as authors.

We then discuss the dedicated indexing software that is used by professional indexers. What is so different about it and why is it useful? Finally, we discuss software used for indexing websites.

1 Fully automatic or semi-automatic indexing software

With automatic indexing software, it is the computer that selects the terms to be included in the index, not the indexer. The term 'automatic indexing software' is somewhat misleading, as the program only generates a concordance list, which is not an index. This means that the software produces an alphabetical list of all the words in the text (excluding a few stop words) – whereas an index is a structured listing, usually alphabetical, of the key concepts of a text.

More sophisticated automatic indexing programs produce selected word lists. Still, at best they produce only the basis for an acceptable index – certainly not the final product. Examples of recent automatic indexing programs are TExtract and IndDoc.

TExtract claims on its website that it is a semi-automatic indexing program. At the same time, however, it claims that you can produce an index with TExtract within several hours. Any professional indexer can tell you, though, that it is impossible to produce an index within several hours, no matter how familiar you are with the text.

Example: In a recently edited index made by somebody else, it quickly became clear that once a decision had been made to include



LERNEN 2.0
Wege in die Zukunft

16. Internationaler Kongress und Fachmesse
für Bildungs- und Informationstechnologie

www.learntec.de
29. – 31. Januar 2008
Messe Karlsruhe

LEARNTEC
Wissen, was kommt.

KMK IDEEN VERBINDEN.
Karlsruhe –
Messen und Kongresse

a specific term in the index, each and every mention of that term was included without any regard for the context. This meant that the term 'body' – which here referred to the female body, as it was a book on women – included several mentions where 'body' had been used as 'body of literature'. Of course, the index also included many insignificant mentions of concepts that should have been left out, but a computer is unable to distinguish between significant and insignificant uses of terms.

2 General-purpose software

You can produce an index using software that is already there on your computer and that you would generally use for other purposes. Often, one's first indexes are produced using Word, for example, but it is also possible to produce indexes with desktop publishing software, spreadsheet programs, or database programs. The advantages are, of course, that you are already familiar with the software and that you won't have to invest in additional software. Word processors are most often used by non-professional indexers. However, the alphabetization creates problems when using these software packages. Word processors sometimes sort text on ASCII character coding. This means that all upper case characters will sort before lower case characters, whereas in indexes sorting should not be case-sensitive. Also, one regularly wants to take small words out of the automatic sorting. For example, one does not want the program to sort on prepositions like 'in', 'for' and 'to' in the subheading, but you may still want to retain them at the start of an entry. The same problem arises with names that start with particles that you do not want to invert, but you would not want to sort on either, e.g. 'al-' in 'al-Qaeda'. This cannot easily be achieved with word processing software. You can change things manually in the sort, of course, and if you only produce an occasional index that may be acceptable. When you have to index regularly, you will soon be looking for better solutions.

There are professional indexers who use Word or a database program for indexing, but these programs are limited in what they can produce in terms of indexing. When professional indexers use Word or Adobe FrameMaker for indexing, they would usually make so-called 'embedded indexes'. This means that the terms selected for the index will be marked with 'tags' in the text. The index is then created by generating a listing of the selected terms. The text and the index are one document.

Producing the 'tags' is rather cumbersome, and embedded indexing is generally more time-consuming than indexing with the dedicated indexing software that is discussed in the section below. In particular, the editing of the index – which in our experience takes up about one-third of the time needed to produce an index – is much more time-consuming with embedded indexing.

Recently, software tools have become available to make embedded indexing less time-consuming and to allow professional indexers to use their dedicated indexing software in conjunction with the Word embedded indexing tool. Examples of these are DEXter and WordEmbed. Both are basically collections of macros that run in Word.

3 Dedicated indexing software

The three major dedicated indexing software packages are: SKY Index, CINDEK and MACREX. All three started out as DOS programs and they all developed Windows versions in the second half of the 1990s. CINDEK is also available for Mac. Demo versions of all three programs are available without time limits.

Basically, all three packages are able to do the same things, although they differ in the way they go about it. CINDEK works with the index card metaphor, MACREX has a very simple input screen that looks like the index itself, and SKY Index uses a database or spreadsheet table metaphor (rows, columns and cells). The interfaces of these programs are very efficient; to put it in Mulvany's words (1994, p. 272): 'Perhaps one of the strongest features of this software is that the indexer can work with the index in sorted order at all times. ... The structure of the index is constantly emerging and visible. The work area context is the index itself.'

All three programs allow you to work in alphabetical order, for example, according to the ISO standard (1996) and the Chicago Manual of Style (2003, pp. 773–784) or in page number order. The latter option is very useful when you have to update an index for a new edition or when several pages of text have been inserted at the last moment. Page numbers can be increased or decreased by a specified amount, and volume numbers can be added, deleted or substituted in a number of ways.

Indexes can be split or joined in various ways. There are several methods of alphabetization available (letter-by-letter and word-by-word), the sort can also be made chronological (very useful for tables

of legal cases in Continental law), or the sort can be forced in other ways (very useful when making an index of Biblical references that has to be in the order of the Bible books).

Dedicated indexing software is excellent in keeping indexes consistent, covering such formal aspects as sorting, punctuation within entries, indents, locator format, and cross-reference consistency, thus enabling the indexer to focus on the index itself.

To highlight the differences between the three most important dedicated indexing software programs, we will present three short reviews of each of them. Each 'review' is written by an experienced user. Within the scope of this article, it is not possible to review these programs in detail; however, the references will give further guidance to those who want to know more about these software tools. There is much more to be learned about each program than can be presented here, and each program has excellent manuals (well-indexed, of course!) and gives support where needed.

Each individual reviewer has made their own choice about which features to highlight from the program. When a particular feature is mentioned in one review and not in the others, it does not necessarily follow that this feature is missing in the other two programs.

3.1 MACREX

MACREX was developed and marketed by Hilary and Drusilla Calvert in the UK. The program has been around for more than twenty years now and it still very much has a 'DOS feeling' to it. For the history of the program, see Beare (2004). Figure 1 shows the simple startup screen that you get once you have started a new index.

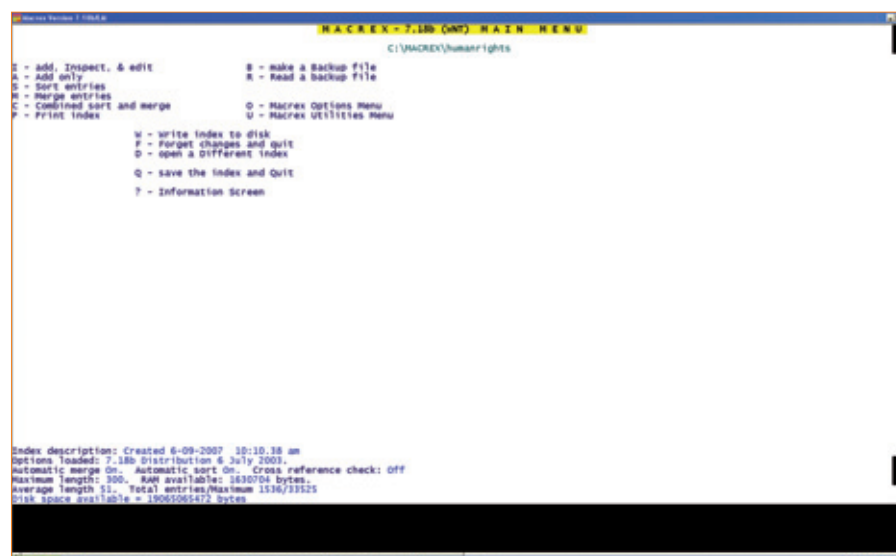


Figure 1: MACREX startup screen

Don't be fooled by the simple appearance, however, because MACREX is a very adaptable program, able to handle many different and complicated indexing situations. It is not very intuitive, though, and it has no pull-down menus. Some of its possibilities are buried in one or two layers of sub-menus and you really have to know how to get there.

The simple appearance of MACREX is going to change soon, as a new release (version 8) is planned for the autumn of 2007. New versions of MACREX always retain the features of previous versions and there is always compatibility with previous versions. Backup files of indexes made with the first versions of MACREX can always be opened in the newer versions. The new features of version 8 will be discussed briefly at the end of this section. First, we will look at the main features of the current MACREX version.

Choosing option 'I' from the startup menu brings you to the screen that allows you

to add and edit entries. Again, the appearance of the screen is very simple. There are no fancy graphics or pop-up boxes for data entry. The screen very much looks like the index itself, as can be seen in Figure 2.

At the bottom of the screen is a line, and below that line you can make your new entries. People who are new to the program may find it hard to remember that you will have to choose 'F4' before you can add new entries. Once you are used to the program, you stop thinking about this. When your entry is ready, it will be inserted at the right place in your index after you have hit the 'Enter' key. When working with MACREX, you do a lot of editing at the input phase of your index. This is because the input screen looks so much like the index itself, you see right away when an entry should be phrased differently, can be merged with a similar entry or when a series of sub-entries does not look good and should be arranged differently.

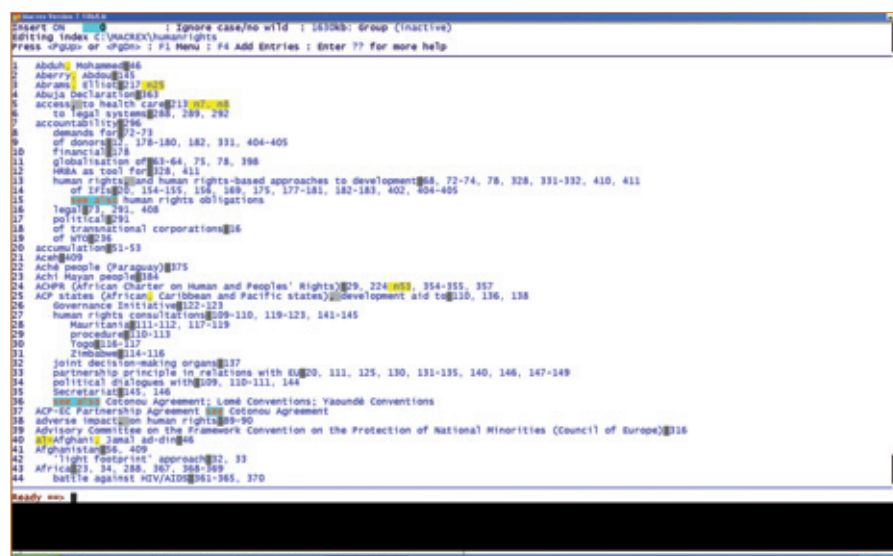


Figure 2: Index input screen

Of course, you don't do all the editing as you go along, there is also editing to be done at the end. For this purpose, the 'group function' is extremely useful. Hitting 'Ctrl H' when in the 'input screen' will get you to the 'group function' screen. Figure 3 shows what the 'group function' screen looks like.

This was created by using one single search term: 'domestic'. All entries containing this term were then selected from the index. This allows you to compare different entries and to check whether your index is consistent. The numbers before the entries indicate the index entry number. Figure 3 shows that the index is not yet entirely consistent, 'domestic state parties' can probably be merged with the much simpler entry 'states' and the entry 'states' needs

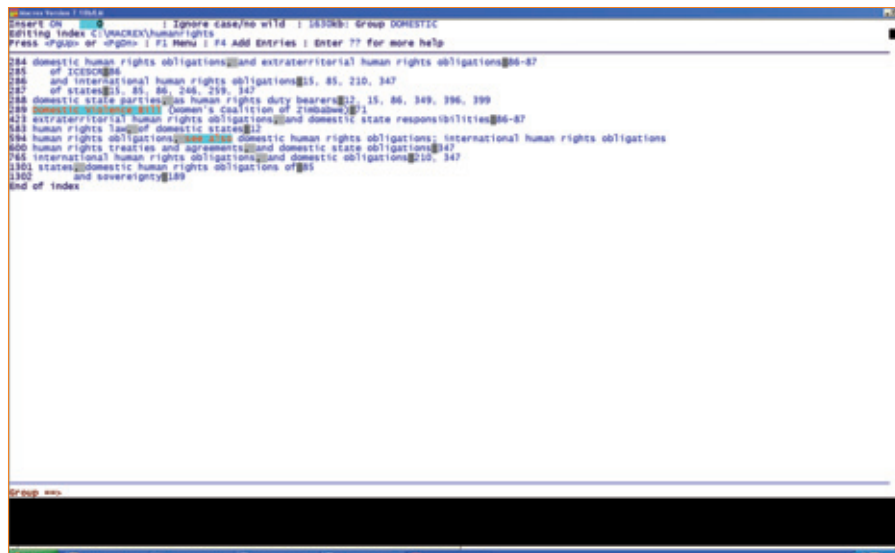


Figure 3: Screen in Group Function mode ('domestic' was used as search term)

additional page locators. This will happen automatically, though, when entry numbers '288' and '1302' are merged.

It is possible to do much more complicated searches in the group mode. You can use 'wildcards' to indicate that you want to search for a specific string of characters only at the beginning or the end of an entry. There are several other types of 'wildcard' as well, and it is also possible to do Boolean searches.

When your index is finished, the output file has to be produced. This is done by choosing the 'Print' option from the startup menu. Again, not a very intuitive choice, and producing the output file is not a very straightforward procedure. Beginners may find this procedure slightly complicated, but again, once you are used to it, it is not much of a problem. There are many layout options to choose from, although in practice one probably only works with one or two of the layout files. It is also possible to have layout

files adapted for you, when publishers have very specific wishes/inhouse styles (e.g. 'OUP' stands for Oxford University Press in Figure 4). To give you an idea of the available choices, Figure 4 shows part of the layout options screen for output.

It is also possible to obtain layout files or other files adapted to your specific needs. These files are easily loaded into the program. For example, special layout files have been produced for use with Unicode fonts (when characters with diacritics are needed) and custom made files have been made for electronic indexing procedures (XML files), that tend to differ from one publisher to another.

As stated at the beginning of this section, MACREX is about to release version 8 of its program. There will be quite a few drastic changes in version 8, and we will therefore briefly describe the new features here. Version 8 will have a much more 'Windows look', with pull-down

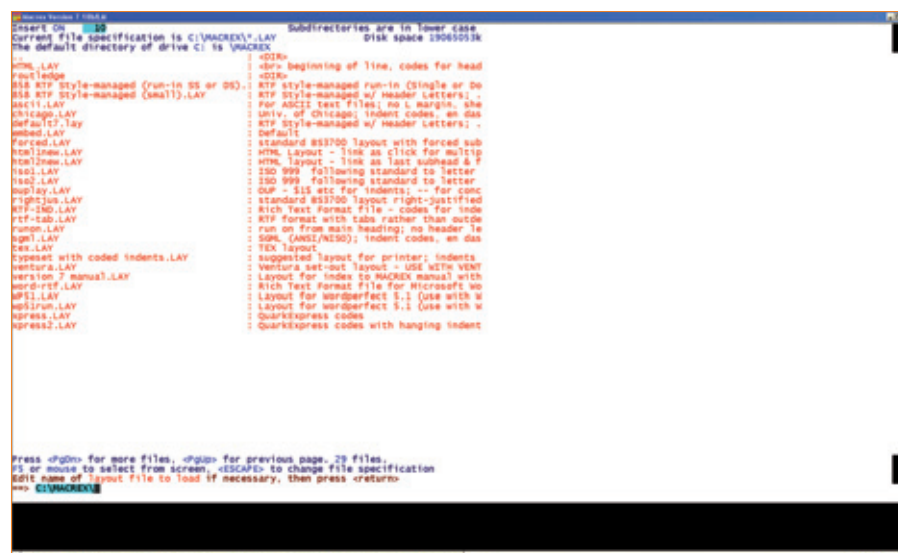


Figure 4: Screen with choices for output formats

menus; however, all the version 7 keystrokes will also be retained. Navigation through the menus will be simpler, with most options available through the main menu, and there will be a context-sensitive help option available at all times. Creating the output file will be much easier, and it will be done in a single step. And there will be an 'autocomplete/table of authorities' feature. This option will also make it possible to work with a controlled vocabulary, as you can easily import lists of terms made in other indexes.

To summarise, when starting to work with MACREX, it is not an easy program to work with. It takes time to get used to the way it works and to find out about all its different options and possibilities. Once you are past this stage, you will find that it is a very reliable and adaptable program that is a joy to work with. Version 8 will probably take away a lot of the irritation that beginners or infrequent users may experience with the program.

3.2 CINDEXTM

CINDEX is produced by the company Indexing Research, founded by Frances Lennie in Rochester (New York), in 1986. The program was originally produced for MS-DOS and underwent continuous development throughout the 1990s, culminating in the current Windows and Mac versions.

Entering records

In contrast to indexing programs that have a spreadsheet-type appearance, records are entered in CINDEXTM in a separate dialogue box with entry line fields for 'Main', 'Sub1', 'Sub2', etc., and 'Page' (Figure 5). The Page Down key then enters the record into the index and clears the box for the next record. This easily used interface means that the record that is currently being worked on is clearly legible and clearly structured (instead of being just one of the screen-wide lines displayed in a screen-long list).

The typing of index records is speeded up by the program's abbreviation and auto-complete features. The user can create abbreviations for phrases that are frequently used; when the abbreviation is typed in the entry box, the program automatically expands it. With auto-completing, the program recognizes when the first letters of a previously entered record are being typed and automatically completes the main heading or subheading line. (All of these features can of course be switched on or off as desired.)

Another much-used facility, as in other programs, is 'flipping', in which the main heading and sub-heading in a record can

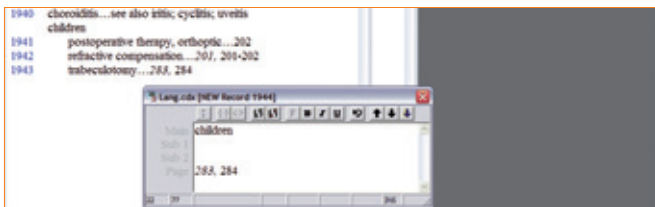


Figure 5: CINDEX entry dialogue box

be switched by pressing a screen button or a key combination. In addition, any heading or subheading from the record entered immediately previously can be copied automatically by pressing a key combination.

Hot keys can be programmed, and 'see' and 'see also' are installed by default as SHIFT F1 and SHIFT F2. The user can also set macros to speed up frequently used sequences of keystrokes.

Displaying the index

At any time during record entry or editing, the index can be displayed in various formats – sorted/unsorted, draft format, or full format with run-in or indented views. Widely flexible sorting and formatting options are available (Figure 6), and a list of subheading prefixes to be ignored during sorting is available and can be edited. Non-alphabetical sorting of subheadings (e.g., for chronological sequences) can be forced by adding hidden coding, and an index can of course also be sorted in page sequence so that additions can be made if there are changes in proof, or for a new edition of a book.

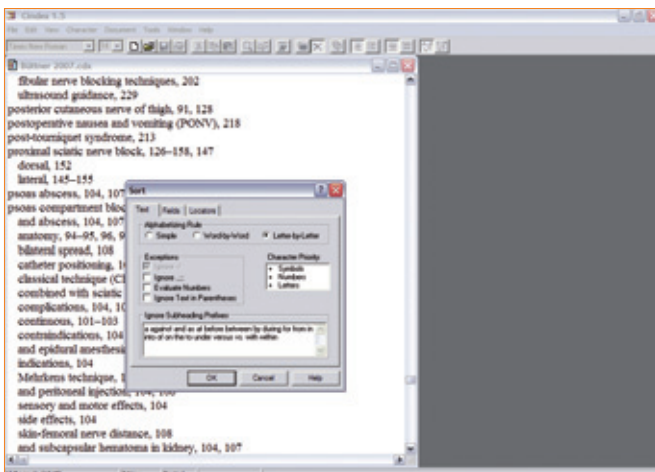


Figure 6: Options for sorting

Editing records

After the entry of records has been completed and the resulting index has been sorted alphabetically (with options for 'simple', 'word-by-word' and 'letter-by-letter'), all records that appear as single entries with a single subheading (e.g., 'knee | anatomy, 159') can be automatically joined to form single main entries ('knee, anatomy, 159').

One of the most useful features of the program during editing is the 'propagate' facility. After sorting, a main entry that needs to be corrected may have a large number of subentries. To avoid the need to edit the main heading in each individual record for all the subentries, activating the 'propagate' function (using a screen button or keyboard shortcut) copies corrections to all of the subentries.

Another important editing function is the facility to group entries when considering the best choice of phrasing or placement for conceptually related terms. Specific related entries can be identified using the 'Find' function and saved as a group, which can be viewed or printed out and edited separately – providing greater clarity and saving time. When editing of the group has been completed, a simple key combination returns the user to the full index, with all of the group edits incorporated into it.

complete an entry session before switching to a separate editing session in which different options are available.

Saving and file format

When work on an index is completed and the file is ready to go to the client, the format for the output file is selected via 'File | Save as ...', with RTF (Word), Quark, and AAP options available (Figure 7).

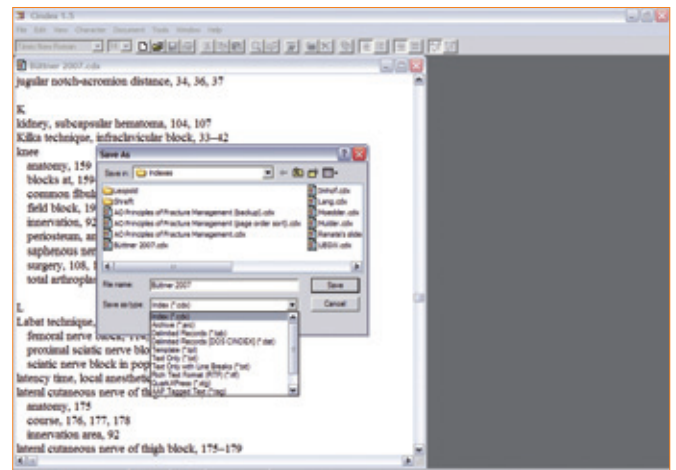


Figure 7: Options for file output

The history and development of the program is covered in Lennie (2005).

3.3 SKY Index

SKY Index (Professional Edition) was first released in 1997 and is under constant development by Kamm Schreiner of SKY Software in the USA.

User interface

The most prominent feature of SKY Index is its unique user interface. As can be seen in Figure 8, there are two main areas. The upper half is called the preview pane, which always shows the developing index in alphabetical order, including main headings, subheadings with indents, cross-references, and locators. The preview pane displays the index as WYSIWYG, including the fonts and text formatting chosen.

The lower half is the data entry grid with its spreadsheet- or database-like look and feel. Here, all the editing of index entries and individual headings takes place, and entries can be added and deleted. The data entry grid consists of columns for main heading, subheading(s), page references (and other types of locators such as volume and chapter numbers) and other details. Cross-references are entered in the page field. The order of entries in the data entry grid can be viewed in various ways – for example, as entered or in page order. This provides an efficient way of working and complements the alphabetical view in the preview pane, so that there are always two views that can be seen at the same time.

Above the preview pane are the menus, main toolbar, sort toolbar as well as acronym and macro toolbars. The main toolbar with its symbols provides more direct access to many of the individual options given in the menus. Some of these individual options and their keyboard shortcuts are the same as in other Windows programs, others are specific to SKY Index.

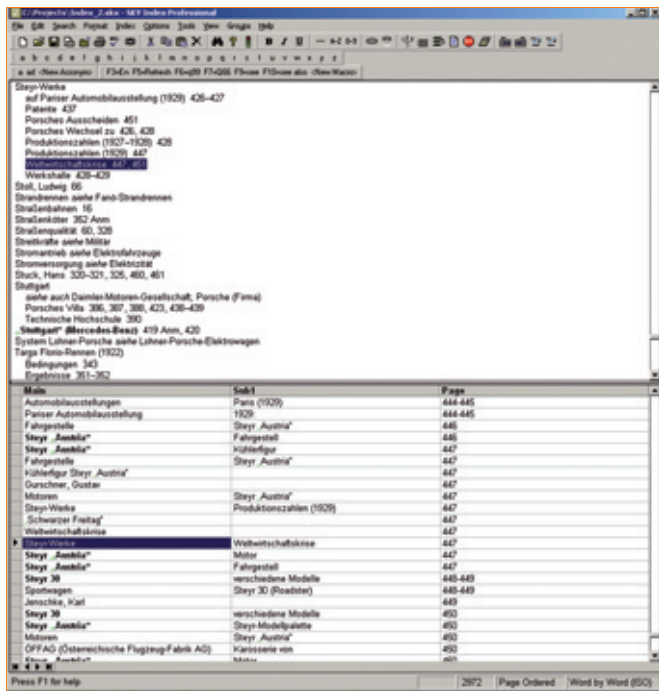


Figure 8: SKY Index main window

Getting around

Navigating in the index is very easy using the mouse or keyboard. Selecting an entry in either the preview pane or data entry grid automatically displays the same entry in the other half. Thus, one can very quickly check out an entry in its alphabetical surroundings and in another view. Moving around the index is also possible by using the sort toolbar (jumping to the corresponding first entry of the selected letter), via a browse modus, by using the scroll bars or arrow keys.

Adding and editing entries

While adding entries, the two features called AutoComplete and AutoRepeat can save a lot of typing. AutoComplete makes use of headings already entered, while AutoRepeat enables certain parts of an entry to be repeated in the next entry. On top of that, by defining acronyms the indexer can efficiently handle long character strings, and macros help to save time triggering actions which are used over and over again.

Entries and individual headings can be manipulated in a large number of ways, including duplicating, flipping, and combining, to name just a few. For example, the following entry can be duplicated and flipped as shown within a split second, while automatically adjusting prepositions to their right locations:

Autobahnen
in Bayern 73

Bayern
Autobahnen in 73

Many index entries can be treated like that at once by simply selecting any entries in the data entry grid.

Another powerful feature makes it possible to isolate (= group), view, and edit large entries with many subheadings, though this is not restricted to large entries alone. Instead, any subset of the index can be grouped and edited, including whole entries, certain words and phrases within entries, and locators. There are many ways of creating groups, which can even be defined and reused.

There is a sophisticated Find & Replace option for altering headings. In addition to that, the Propagate Edits function can alter many headings with identical text at once. Another useful option is converting cross-references into entries (double posting).

Sorting

Among the many sorting methods available are the following ones. Entries beginning with special characters can be force-sorted using Hide or Ignore text options; this formatting is displayed in colour and can easily be seen. Sorting of Roman numerals and dates is possible. Prepositions can be ignored when sorting subheadings, as in:

Autobahnen
in Bayern 73
mit Gebühren 41, 92-95
Glatteis auf 127-129

Quality control

The Error Scan tool can detect important errors, such as cross-reference inaccuracies, improperly formed entries, orphan subheadings, and too many locators behind headings. All detected errors are

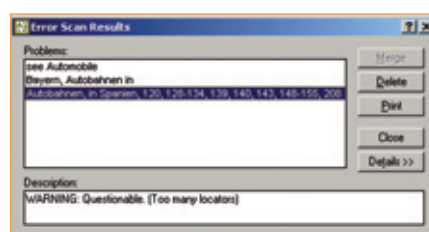


Figure 9: Error Scan Results

listed in a dialog box (Figure 9) and can be fixed via the data entry grid.

At any time it is possible to produce an output of the whole index or parts of it. Among the output formats available are various RTF variants, HTML, and Quark.

Overall, the program can be used in a very intuitive way, enabling even beginners to enter entries instantly, while experts benefit from the many useful features and configurations they can choose from.

A comparison of CINDEXT and SKY Index has been published by Michael Wyatt (2001).

4 Website indexing software

There are different ways in which websites can be searched or made searchable. Very often one will find some sort of search engine as a tool to search a particular website, although there are other tools available as well (such as navigation menus, site maps, searchable databases, etc.). A back-of-book style index is another way of searching a website. On websites that have this feature, one would typically find an A-Z alphabet somewhere on the home page of the website. Clicking on one of the letters gives you access to the index. Two articles on website indexing itself, by Heather Hedden and Glenda Browne, can be found on pages 433 and 437 in this issue.

Specialized software for web indexing is available, but it is not very well known. Not very many professional indexers do website indexing, although this may change as more and more courses and workshops on website indexing are on offer. Professional web designers in general know very little about indexing and the software tools available. Using dedicated indexing software, it is possible to make indexes with HTML output. In combination with an HTML editor, one can also do website indexing, MACREX being the most flexible program in this respect. However, it is easier to use specialized website indexing software, and we will therefore briefly look into the two best-known website indexing software tools available: XRefHT and HTML Indexer.

4.1 XRefHT

XRefHT (commonly pronounced 'Shrefht') is a freeware program developed by Tim Craven, professor of the Faculty of Information & Media Studies at the University of Western Ontario. The program can be downloaded from Tim Craven's faculty website.

The XRefHT main window (Figure 10) is relatively simple and consists of the three columns Heading, Subheading, and URL. It is possible to load these columns with the titles of HTML files, anchor names and headings <h1> to <h6> from within HTML files, and also keywords contained as meta data in the head of such files. This can be done with local files on one's computer as well as with single web pages or even a whole website. XRefHT extracts the respective data from the files selected. The URL column represents the locator field; it accommodates the file name (of HTML, PHP, and other file types) as well as subfolder and anchor name, if any.

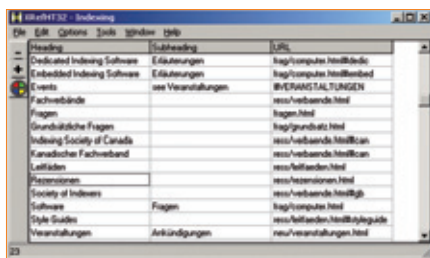


Figure 10: XRefHT main window

Once the columns have been populated, the index entries can be edited – for example, rephrasing headings and adding subheadings, as well as duplicating, inserting, or deleting entries. It is possible to sort on headings or URLs at any time in order to keep track of the developing index. It is also possible to preview the HTML output of the index in a browser window.

See and See also cross-references can be added in the Heading or Subheading columns by writing 'see' or 'see also', followed by the target entry; the target anchor name has to be entered in the URL field. There is an alternative way of adding cross-references to an index when using XRefHT in conjunction with the thesaurus software TheW32, another freeware program available from Tim Craven.

There are a number of other features, such as the option of adding anchor names to HTML files, a spell checker, and XRefHT's own editing and web browser windows. For a freeware program, it does have a surprisingly large number of features, although the editing functionality is not as efficient as that of commercial indexing software. Also, the level of documentation is somewhat inadequate. However, there is a good description of the program by Heather Hedden (*Indexing Specialties: Web Sites*, pp. 61–77) and a whole book about it by James Lamb (2006).

4.2 HTML Indexer

HTML Indexer is a commercially available software tool developed by David Brown.

HTML Indexer's capabilities are similar to those of XRefHT, but it has several advantages over XRefHT. A free demo version is available. The advantage of using commercial software over freeware is, of course, that one is entitled to receive technical support.

In HTML Indexer, you start an index by starting a new project and adding the website files that you want to index to your project. The entire site can be extracted at once. Figure 11 shows the 'project tree pane' that you see on the screen once you have added the website files to your project.

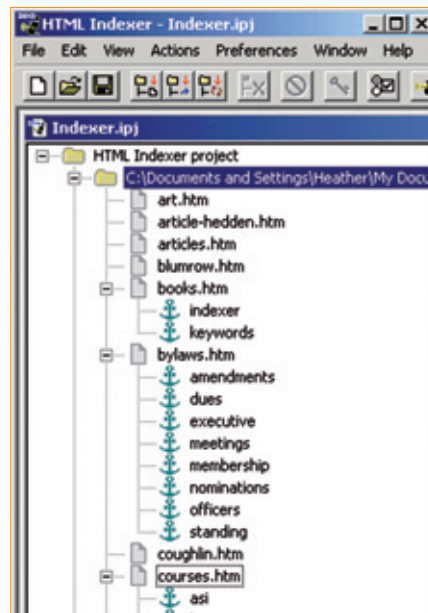


Figure 11: Project tree pane

The subfolders and the files are arranged in alphabetical order by title.

If there are subfolders, they will have plus signs next to them, which you can click once to expand the contents of the subfolder and see the list of files within.

Some of these files in turn have plus signs next to them. This is to indicate that the file has internal anchors. If you click once on the plus sign, the file expands to show the anchors within the file, as named, with an anchor icon.

Any item that is selected in the left pane by clicking once on its icon, rather than on the plus/minus sign, will then call up a list of URLs in the right pane.

It is in this left Project Tree pane that you will select the files and anchors for indexing. You can also select them from here to view in your browser when deciding how to index. (Source: Heather Hedden, *Creating Website Indexes*, 2007)

URLs of both web page file names and anchor targets are extracted into the

indexing program in a single step. Subfolders are also extracted. Non-HTML files can also be indexed. These files have to be added to your project by using the 'Actions menu'. You then select files and anchors for indexing. This brings you to the 'index entry pane', illustrated in Figure 12.

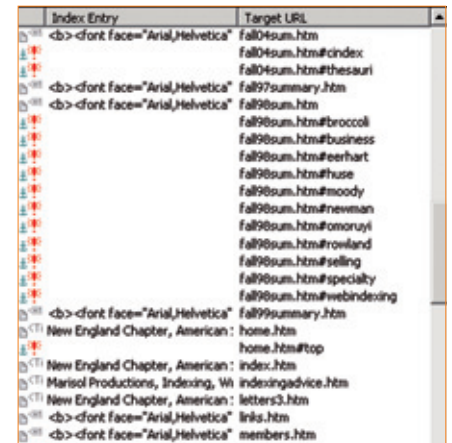


Figure 12: Index entry pane

In the right pane there are three columns. In the first column are icons indicating the source of the URL, a web page, or an anchor. The second column is for the Index Entry, which may have text that is automatically generated with something that needs to be replaced. The third column is for the Target URL, the web page name or the page name with an anchor. If there are subfolders, the folder paths will also be included in the URL.

If there is no automatically generated text for the Index Entry, which is usually the case with anchors, then a red exclamation mark appears.

There is no column for subentries. Subentries are entered as part of the main entries. (Source: Heather Hedden, *Creating Website Indexes*, 2007)

According to Heather Hedden (2006, p. 36), creating additional index entries is easier and more sophisticated than in XRefHT. And unlike XRefHT, the default text that is automatically entered is not the web page title, but the contents between the first pair of heading tags on the page. If there are no heading tags, then the page title is used. Cross-references can be added at any time. There is a special 'create cross-reference' box that allows you to make a cross-reference and also have it hyperlinked.

There are many index styles you can choose by using the 'preferences' menu. You can also use CSS (cascading style sheets). HTML Indexer comes with a default style sheet, and you can use this or create your own. Mike Unwalla (2006) has shown that, for people who know

about HTML coding, it is fairly easy to adapt HTML Indexer in various ways by including user-defined HTML code.

HTML Indexer is different from other web indexing tools in that it has a method for self-maintenance. It actually stores the index entries on the page indexed by adding a comment block, which cannot be seen by visitors to the website. There is a corresponding comment block on the index page. This ensures that no index entries are created to locations in the website that do not exist, or no longer exist.

5 Conclusion

In this article we have looked at software that is used for indexing purposes, noting that automatic or semi-automatic indexing software does not create acceptable indexes. Automatic indexing can at best provide a starting-point for further indexing. It is questionable whether this is a better starting-point than just reading the text, because only the context can help you determine whether a term is being discussed or whether it is just a passing mention that does not deserve to be included in the index.

For someone who needs to produce an index only occasionally, using general-purpose software is an acceptable solution. Each program has its drawbacks, because they were, after all, developed for other purposes. For example, these programs lack sophistication in alphabetic sorting or alternative sorting orders (page order, chronological) and are time-consuming during the editing phase, as the indexer does not have an overview of the index.

Dedicated indexing software is therefore nearly always used by professional indexers. These programs are far superior in the way they handle data entry, manipulation of data and editing of data. They are very flexible in the output formats they can produce and are adaptable to different indexing situations, such as traditional back-of-the-book indexing, embedded indexing and website indexing. The three most widely used dedicated indexing programs are reviewed here. We found that their orientation and appearances are very different, but that the features (e.g. 'group function') and final output are often remarkably similar. The program of choice is therefore mainly a matter of personal taste.

For website indexing, other software solutions are available as well. And, as this is a fairly new indexing speciality, it

may very well be that more and more sophisticated products will become available in this field.

We hope we have shown that indexing involves intellectual effort that is greatly enhanced by the use of software, but cannot be replaced by software (i.e., automatic indexing software).

References

- Beare, Geraldine: MACREX – a History. *The Indexer* 24(2004)1, pp. 18–20.
Chicago Manual of Style. 15th ed. Chicago: University of Chicago Press, 2003.
CINDEX website: www.indexres.com/home.php [1st October 2007].
DEXter website: www.editorium.com/dexter.htm [1st October 2007].
Hedden, Heather: Software for HTML indexing: a comparative review. *The Indexer* 25(2006)1, pp.32–37.
Hedden, Heather: Creating Website Indexes. NIN workshop handout. Driebergen, August 2007.
Hedden, Heather: Indexing Specialities: Web Sites. Medford, NJ: Information Today, 2007.
HTML Indexer website: www.html-indexer.com [1st October 2007].
IndDoc web page: www-lipn.univ-paris13.fr/RCLN/outils/IndDoc/presentation.htm [1st October 2007].
ISO 999: Information and documentation – Guidelines for the content, organization and presentation of indexes. Second edition, 1996-11-15.
MACREX website: www.macrex.com [1st October 2007].
Lamb, James: Website Indexes: Visitors to Content in Two Clicks. Colchester: James Lamb, 2006.
Lennie, Frances S.: History and development of CINDEX. *The Indexer* 24(2005)3, pp. 135–137.
Mulvany, Nancy C.: Indexing Books. Chicago: University of Chicago Press, 1994.
Mulvany, Nancy C.: Software tools for indexing: revisited. *The Indexer* 21(1999)4, pp. 160–163.
Schroeder, Sandi (Ed.): Software for Indexing. Medford, NJ: Information Today, 2003.
SKY Index website: www.sky-software.com [1st October 2007].
TExtract website: www.texyz.com/textract/ [1st October 2007].
Tulic, Martin: Software for Indexing. Website article, 2004: www.anindexer.com/about/sw/sw_index.html [1st October 2007].
Unwalla, Mike: Web indexing: extending the functionality of HTML Indexer. *The Indexer* 25(2006)2, pp. 128–130.
WordEmbed website: www.jalamb.com/wordembed.html [1st October 2007].
Wyatt, Michael: CINDEX vs Sky Index. Website article, 2001: www.aussi.org/resources/software/review.htm [1st October 2007].
XRefHT download site: <http://publish.uwo.ca/~craven/freeware.htm> [1st October 2007].

Index, Indexing, Software,
Evaluation, Overview

Register, Indexieren, Software,
Übersichtsbericht, Bewertung

THE AUTHORS

Caroline Diepeveen, MA



has university degrees in political science and international relations. She worked on various research projects and worked in development cooperation, before relocating to Scotland and taking up indexing. In 2000, she moved back to the Netherlands, where she continued her indexing career. Caroline has been an accredited indexer with the British Society of Indexers since 1998. In 2004, Caroline and others started NIN, the Netherlands Indexing Network.

Brandenburglaan 39
4333 BX Middelburg
The Netherlands
cdiep@zeelandnet.nl
www.cdiep-indexing.com

Dipl.-Dok. Jochen Fassbender



is a freelance indexer and information architect with a background in information science, based in Bremen, Germany. He is co-founder of the German Network of Indexers (Deutsches Netzwerk der Indexer – DNI) and presently its main organizer. He is also a member of the Information Architecture Institute. Since 1991 and prior to becoming a freelancer he worked in several long-term projects (indexing as well as thesaurus development) for various institutions and companies.

Indexetera – Indexing et cetera
Bremen, Germany
jf@indexetera.de
www.indexetera.de

Dr. Michael Robertson



studied English at the Universities of Stirling and York (UK). He worked in university research and science publishing for several years, and has been a freelance editor, indexer and translator, based in Augsburg (Germany), since 1991.

Editorial and Translation Services
Augsburg, Germany
editchoice@yahoo.com
www.editchoice.com