

## 32 - INDEXING OF ELECTRONIC DOCUMENTS

- 32.0 Introduction
- 32.1 Method one: Multiple references per entry
- 32.2 Hypertext markup menu
- 32.3 Making index entries
- 32.4 Checking cross-references
- 32.5 Making the final HTML index file
- 32.6 Alternative method: One reference per entry
- 32.7 Using an HTML source file in the Netscape editor
- 32.8 Notes
- 32.9 MACREX menu options relevant to hyperlinking
- 32.10 Checking the file in your internet browser

### 32.0 INTRODUCTION

We have introduced optional features into MACREX to produce the text necessary for hypertext links in place of a page reference. There are two separate methods: the first will support any number of references; the second supports a single reference to a particular entry. The examples following use HTML as an example, but the system should be adaptable to other markup systems.

### 32.1 METHOD ONE: MULTIPLE REFERENCES PER ENTRY

This example uses an HTML source file in a text editor. We use Textpad, which is an excellent shareware product available from <http://www.textpad.com/>, but you can equally well use the Notepad (but note that file size on Notepad is limited to 30K) or the Wordpad text editors provided in the Accessories program group on all versions of Windows from 95 onwards. We no longer recommend that you use Word, in spite of the fact that Word has the facility to save HTML documents because each file is "bulked" up to such an extent that the speed of access is considerably slowed down. However, if you need to use Word, we have instructions which we can send out to interested people.

1. Open the HTML document in Textpad or your chosen text editor.
2. Open a new index in MACREX either from the Macrex-7 program group or from the MSDOS prompt.
3. Go to the **Options menu**
  - a. Set option **R - Page number increment** to a suitable number (1 will do, but if you set it to a number higher than one the numbering will leave gaps to fit new entries in between the ones you create initially (e.g. if you set the increment to 5, your numbering will go 5, 10, 15, 20 etc).
  - b. Set option **S - Current page number** to a suitable starting value (1 will do). The number used to form your first hypertext locator will be the value set here plus the page number increment.
  - c. Select option 7 - Hypertext markup features and you will get a new screen:

## 32.2 HYPERTEXT MARKUP MENU

### MACREX HYPERTEXT MARKUP MENU

- A - Add hypertext markings? No
- B - Text before page number `<a href="#"`
- C - Text after page number `>[*]</a>`
- D - Clipboard text before `<a NAME="`
- E - Clipboard text after `></a>`
- F - Numerical field width 4
- G - Text before xref link `{<a href="#"`
- H - Text after xref link `>}`
- I - Text after xref text `{</a>}`
- J - Text before xref target `{<a NAME="#"`
- K - Text after xref target `></a>`

Change **A - Add hypertext markings?** to **Yes**. Leave the other options unchanged. Save this setting as the default by pressing **<escape>** and choosing "Save in index directory (indexes in this directory only)".

## 32.3 MAKING INDEX ENTRIES

4. Go back to the Main Menu (press **<enter>** twice) and then press either
  - i) **A** to go to the **Add only** screen
  - or
  - ii) **I** to get to the **add, Inspect, & edit** screen and press F4 to add entries.
5. Input the text of your first index entry. You can do this by either
  - i. Typing whatever you want in the usual way
  - or
  - ii. Swapping to the source document (**ALT-TAB**) in Textpad (or whatever editor you are using), highlighting some suitable text and then using **CTRL-C** to copy it to the clipboard, and then swapping back to MACREX (**ALT-TAB** again) and using **CTRL-V** to insert it.
6. Once the text of your entry (with whatever headings, subheadings you want) is completed, press **F9<enter>**. In our example, if the text was **sorting, Greek characters**, the first entry will appear as follows:

```
sorting, Greek characters {<a href="#000}2{ ">[*]</a>}
```

This works as follows: In the middle we have **2** which is the number on which the locator is built. Everything before and after is enclosed in **{ }** so that MACREX will be quite happy to treat it as a normal page reference. The `<a href="#"` is derived from option **B - Text before page number** on the MACREX HYPERTEXT MARKUP MENU and is the first part of the HTML syntax for a hypertext link to a label which will be **0002**. The three zeros are inserted by MACREX because the setting for **F - Numerical field width** is **4**, making the total field width 4 characters (you can change this to whatever you like). The **[\*]** is the "click text" - what the user clicks on to locate the entry (you may wish to change this to **[a]** for the first reference, **[b]** for the second, etc.). The rest of the reference is the closing sequence for the HTML hypertext link.

*Note that the defaults are designed assuming you want to incorporate the HTML file produced by MACREX into the file that you are indexing at the end of the job. If you want the HTML file containing the index to be a **separate file**, possibly referencing locations in many different source files, you can edit **B - Text before page number** to include the name of the file currently being indexed. For example if you were editing a file called **desc.htm** you would change option **B** to `<a href="desc.htm#`.*

You now have your first entry. To make it possible to use F11, F2, F3 and F5, you now need to go to the Main Menu and then back to the **I - Inspect, add and edit** screen. If you want to make other entries (synonyms, double entries, etc.) to refer to the text you have chosen, you can type the new text and press F11. It is only the use of F9 that causes the number to increment.

7. Use **ALT-TAB** to swap back to Textpad (or the editor you are using) and the HTML source. Make sure the cursor is positioned exactly where you want the index entry to take the reader and press **CTRL-V**. The text

```
<a NAME="0002"></a>
```

will appear in the text at this point. This has been created by MACREX and will be automatically have been placed on the Windows clipboard. It is formed in a similar way to the locator above, but uses the text in **D - Clipboard text before** at the beginning and **E - Clipboard text after** at the end. The default text used for these options leads to the formation of a tag for the target that is guaranteed to be the same as that used for the reference in the index.

8. Back in your index, you can edit the text of the entry in MACREX if you wish, providing you don't change anything in the locator other than the text within square brackets. As mentioned in (6) above, if you want to create more entries at any time to link to the same point in the text you can use **F11** to duplicate the hyperlink of the "logical previous entry". F2 (half flip) and F3 (full flip) can also be used.
- 9.. For the **second and subsequent entries** use the following pattern:
  - a) **ALT-TAB** to go to the HTML document in your text editor (if it's not already on the screen in front of you) and choose the next entry.
  - b) **Either block the text** you want and press **CTRL-C**, then **ALT-TAB** back to the index and press **CTRL-V** to insert the text of the new entry *or* **ALT-TAB** straight back to the index and **type in the text** of the entry
  - c) Press **F9<enter>**.
  - d) **ALT-TAB** back to the document.
  - e) Position the cursor where you want the next link (if you haven't done so already) and press **CTRL-V<enter>**.

## 32.4 CHECKING CROSS-REFERENCES

After you have finished the index completely (don't do this at an earlier stage or half-way through the index), back it up (option B from the Main Menu) and go to the **Merge** menu; press **C** to change options. Change option **1 -Check cross-references?** to **Yes** and check your cross-references, and make any necessary changes. When you are satisfied that your cross-referencing is accurate, go back to the Merge options menu and change option **0 - Insert hypertext xref links** to **Yes**. (This will default back to No immediately after you have merged the index even if you press <escape> at this point to save the layout; this is to avoid unwanted extra coding being inserted.) At the main Merge menu choose **S - start merging**. This will use the cross reference term as the link (from the see also reference) and the target of the link in the target entry.

For example

Afrikaners, ^see also^ Boers

and

Boers 181-182

will become

Afrikaners, see also {<a href="#BOERS">}Boers{</a>}

and

{<a NAME="#BOERS"></a>}Boers 181-182

There are five options on the **Hypertext Markup Menu** (accessed from the **Options menu** option 7) to let you define the text used for the hypertext markup - in this example HTML - for cross-references. Where these bits of text are used is indicated by the bold text above. The options on the special hypertext menu with their HTML defaults are:

- G - Text before xref link {<a href="#"
- H - Text after xref link ">}
- I - Text after xref text {</a>}
- J - Text before xref target {<a NAME="#"
- K - Text after xref target "></a>}

Note that the text use for the links ("BOERS" in the example above) is derived from the cross reference term. If the cross reference term is very long (for example

SA (Sturm Abteilung{,} stormtroopers{,} brownshirts)

the text used for the link will also be long; in the current example

<a href="#"SASTURMABTEILUNGstormtroopersBROWNSHIRTS">

This means

- (i) that you need to make sure that there is **plenty of spare record length** before running the merge with the option to insert the links and
- (ii) that the Print program has to know that these pieces of text are not shown on the screen, otherwise it will produce inappropriate line breaks.

## 32.5 MAKING THE FINAL HTML INDEX FILE

Go to the print subprogram, load **HTML.MWP** and **HTML1.LAY** and choose **W - make a Word processor file** (changes and new options necessary are listed at the end of the documentation). The resulting file can be added to the end of your HTML document and will form the index.

## 32.6 ALTERNATIVE METHOD: ONE REFERENCE PER ENTRY

Follow 1-3 in example 1 above, but at the MACREX HYPERTEXT MARKUP MENU change option **A - Add hypertext markings?** to **Yes** and change option **C - Text after page number** to "> (remove "[\*]</a>") or just load the file **HTML2.HTX**.

Follow 4 - 10 in example 1 above.

Go to the print subprogram, load **HTML.MWP** and **HTML2.LAY** and choose **W - make a Word processor file** - (changes and new options necessary are listed at the end of the documentation). This file can be added to the end of your HTML document and will form the index.

## 32.7 USING AN HTML SOURCE FILE IN THE NETSCAPE HTML EDITOR

Follow 1-3 in example 1 above, but at the MACREX HYPERTEXT MARKUP MENU change option **A - Add hypertext markings?** to **Yes** and then call up option **D - Clipboard text before** and delete the default text by pressing <escape> and then call up **E - Clipboard text after** and do the same (delete the default text by pressing <escape>. This is because Netscape composer will generate the HTML coding to go with the target by itself and doesn't need MACREX to do it.

1. Do 4-6 as in example 1 above.
2. Use <ALT>TAB to swap back to Netscape Composer. Make sure the cursor is positioned where exactly where you want the index entry to take the reader and then select Insert->Target (<ALT>IA if you don't like the mouse). When the target

properties dialogue box is displayed press <CTRL>V, which will automatically paste the target name 0002 then press OK.

3. Repeat swapping between MACREX and the source editor until you have done the index.
4. Create the cross-reference links as in 10 in example 1 above.
5. Go to the print subprogram, load HTML.MWP and HTML2.LAY and choose W - make a Word processor file - (changes and new options necessary are listed at the end of the documentation). This file can be added to the end of your HTML document and will form the index.

## 32.8 CHECKING THE FILE IN YOUR INTERNET BROWSER

After you have made your index file, and, if needed, added it to the end of the document, you will want to check that your links work and that the file looks right. With all browsers (Internet Explorer, Netscape, Mozilla, etc.) the system is the same. Open the browser as usual (either on- or off-line), but press <escape> when it appears on the screen. Some browsers will ask to confirm that you are working offline. Then type the path and file name of the document in the box at the top of the screen (the place where you would normally type the URL that you want), e.g.

```
C:\macnt\test.htm
```

The file should appear on screen. Press CTRL-END to get to the end of the text (if that is where you have put your index) and click on some or all of the links to make sure that they point to the correct text.

## 32.9 NOTES

- By changing the options on the HYPERTEXT MARKUP MENU it should be possible to accommodate other coding systems, eg SGML, XML.
- The repeated operations in MACREX (for example the sequence of commands <CTRL>V, F9, <ENTER>) can be made into a MACRO.
- Depending on whether the text or HTML editor has a macro facility, some of the repeated tasks done in the editor may be partly automated.
- This system does not require that you index sequentially. The links will still work even if you jump around. However for editing and checking it will be easier if they are more or less sequential.
- The lowest level of subheading in the entry forms the link.
- All the options in the HYPERTEXT MARKUP MENU are automatically saved in the .INF file with each individual index.
- The settings of the HYPERTEXT MARKUP MENU can be saved as a default or a named file, with the extension .HTX. We have at present two files which can be loaded, both described above, HTML1.HTX and HTML2.HTX. In addition, for Word users, we have two dedicated files.
- The algorithm used to generate sequential unique tags which currently produces numbers with optional leading zeros could, if required, be changed to produce (for example) numbers with a different base (base 36 will produce combinations of all 26 letters of the alphabet and the digits 0-9) or roman numerals.

## 32.10 MACREX MENU OPTIONS RELEVANT TO HYPERLINKING

(options relevant to hyperlinking are in bold)

### OPTIONS MENU

M A C R E X   O P T I O N S	
A - List on printer? No	<b>S - Current page number 14</b>
B - List in a file? No	T - Set video type Windows NT Console
C - Name of list file E:\hyper.LST	U - Intelligent flip? Yes
D - Printer number 1	V - Maximum number of keywords? 100
E - Capitalize flips? No	W - Type of screen list Wrap around
F - Check on deleted entries? Yes	X - Change index description
G - Backspace deletes? Yes	Y - Minutes between automatic saves 0
H - Recognize roman numbers? No	Z - Version 5 page number checking? No
I - Insert on? Yes	1 - Number of heading levels (fields) 0
J - Autosort On	2 - Keep scroll buffer? Yes
K - Automerge On	3 - Set Codepage 850
L - Track added entries? Yes	4 - Printer lines per inch 6
M - Set Colours	5 - Printer characters per inch 12
N - Screen width 80 (77)	6 - Check for bad page ranges? Yes
O - Screen height 25 (27)	<b>7 - Hypertext markup features</b>
P - Change EGA 25/35/43 line mode	8 - Version 6 page number checking? No
Q - Beep with messages Yes	9 - Macro options
<b>R - Page number increment 1</b>	0 - Macrex options menu 2

Pressing 7 from the Options Menu will produce

MACREX HYPERTEXT MARKUP MENU	
<b>A - Add hypertext markings? Yes</b>	
<b>B - Text before page number &lt;a href="#"</b>	
<b>C - Text after page number "&gt;</b>	
<b>D - Clipboard text before &lt;a NAME="</b>	
<b>E - Clipboard text after "&gt;&lt;/a&gt;</b>	
<b>F - Numerical field width 10</b>	
<b>G - Text before xref link {&lt;a href="#"</b>	
<b>H - Text after xref link "&gt;}</b>	
<b>I - Text after xref text {&lt;/a&gt;}</b>	
<b>J - Text before xref target {&lt;a NAME="#"</b>	
<b>K - Text after xref target "&gt;&lt;/a&gt;}</b>	

**MERGE OPTIONS MENU**

## MACREX MERGE OPTIONS MENU

- |  |  |
|--|--|
| A - First letter to be merged A          | S - "See also" separator ;                 |
| B - Last letter to be merged Z           | T - Text before "See also" ,               |
| C - Messages on printer? No              | U - Posn of roman page refs To start       |
| D - Merge to a new file? Yes             | V - Merge unsorted batches? No             |
| E - Compress page references? No         | W - Posn of bold page refs? Float after    |
| F - Squash page references? Yes          | X - Italic/underlined refs? Float after    |
| G - Type of squash Unsquash              | Y - Check text in { }? Yes                 |
| H - Consume internal references? No      | Z - Text for "see" references ^see^        |
| I - Insert second volume number? No      | 1 - Check cross-references? Yes            |
| J - Page refs to renumber None           | 2 - Tag bad cross-references? No           |
| K - First page to renumber 0             | 3 - Cross-ref/page range tag text !!       |
| L - Last page to renumber 0              | 4 - Insert a volume number No              |
| M - Change renumbered pages by 0         | 5 - Vol number to insert                   |
| N - {Annotated }page references Float a6 | a6 - Edit capitalization table             |
| O - Ignore case when merging? Yes        | 7 - Sort {-} as negative number No         |
| P - Placement of ^see also^ Separate     | 8 - Tag bad page ranges No                 |
| Q - Re-merge ^see also^ ? No             | 9 - Insert missing ^ and \ in vols Yes     |
| R - "See also" text ^see also^           | <b>0 - Insert hypertext xref links Yes</b> |

**PRINT LAYOUT OPTIONS MENU 2**

## MACREX PRINTING LAYOUT OPTIONS - MENU 2

- |  |  |
|--|--|
| A - Right-justified filler <space>                 | S - Right-justified page margin 75                   |
| B - Text before header letter                      | T - Flush page numbers to right? Yes                 |
| C - Text after header letter                       | U - Max margin without R flush 75                    |
| D - Minimum run-on level 0                         | <b>V - Punct between text &amp; pages &lt;/a&gt;</b> |
| E - Volume number separator .                      | W - Punct before "see also"                          |
| F - Repeated volume numbers? No                    | X - "see also + <pages>" as subhead? No              |
| G - Page number separator ,                        | Y - Set up table for end of line codes               |
| H - Page number concatenator -                     | Z - Define line break characters -,                  |
| I - Hard Hyphen in page numbers? No                | 1 - Text on beginning of entry                       |
| J - More volume/page options                       | 2 - Formfeed characters                              |
| K - Blank line before new heading? No              | <b>3 - Text when finished printing &lt;/html&gt;</b> |
| L - Text on end of entry                           | 4 - Respect CaSe of headings ? No                    |
| M - Use code table for indents? No                 | 5 - Swap ", at end of heading? No                    |
| N - Set up code table for indents                  | 6 - Swap ', at end of heading? No                    |
| O - Line feed for printer                          | 7 - Swap ^, at end of heading? No                    |
| <b>P - Line feed for word processor &lt;br&gt;</b> | 8 - Swap \, at end of heading? No                    |
| Q - Enter left margin code                         | 9 - MENU 3   |
| R - Force first subhead to run-on No               | 0 - MENU 1   |

**PRINT LAYOUT OPTIONS MENU 3**

## MACREX PRINTING LAYOUT OPTIONS - MENU 3

- A - Text at end of cross-refs
- B - Print group listing? No
- C - Text for group listing
- D - Merge page numbers? Yes
- E - Text between x-refs & run-on subhead :
- F - Reset Layout
- G - Page refs before last subheading Yes**
- H - Character to use for space for indents \_**
- I - HTML Codes in WP file? Yes**
- J - Menu 2

**Option G** in the above menu (layout menu 3) allows you to use something other than a space to indent headings, and was introduced so that you can use **&nbsp;** for a space on the HTML layout without having to fill in all the code table for indents. There is also a problem with the code table if the "space" text is very long because only 80 characters are allowed for each indent, limiting the subheading level to about 3. This can be overcome by not using the code table and setting the appropriate text for indents.

If **option H** on the above menu (layout menu 3) is set to **Yes** the page reference is moved in front of the last subheading.

## OPTIONAL PRINTING FEATURES MENU

### MACREX OPTIONAL PRINTING FEATURES

- A - Install printer Underline/Italic codes (^...^)
- B - Install codes for Bold/Alternative typeface (\...\\)
- C - Change or delete printer replacement codes
- D - Change or delete word processor replacement codes
- E - Add new replacement codes (printer and word processor)
- F - Select type of word processor file
- G - Define Custom word processor codes**
- H - Install codes to initialize the Printer
- I - Install codes to initialize the word processor**
- J - Remove all printer and word processor codes
- K - Set default extension for word processor file name htm**
- L - Name of printer initialization file

#### Option G changes:

Bold: `<60>b<62>` (bold on) and `<60>\b<62>` (bold off)

Italic: `<60>i<62>` (italic on) and `<60>\i<62>` (italic off)

**Option I** is set to `<60>html<62>`