2 Converting word processor files

What's covered in this chapter

You can convert word processor and spreadsheet files to HTML using HoTMetaL PRO. This chapter describes preparing and converting documents that have been prepared in a word processing application. Topics include:

- Converting a single file by opening it in HoTMetaL PRO Editor.
- Converting multiple files at once in HoTMetaL PRO Information Manager.
- Preparing files for conversion.
- Setting conversion options.

Existing files prepared in common word processing, spreadsheet, or text editing applications are easily converted to HTML, using the HoTMetaL PRO Editor or Information Manager.

If you selected Compact Install when you installed HoTMetaL PRO, the conversion filters were not installed.

In this case, the Information Manager commands Convert Documents... and Document Conversion Options... will not be available, and the Editor Open dialog will not contain a list of word processor file extensions.

To add the filters, run the HoTMetaL PRO installation again:

- Run *setup.exe* from the *hmpro4* folder on the CD.
- Choose Custom Install
- Add filters for the formats you wish to convert.

Converting existing files

You can use Open... in the File menu in HoTMetaL PRO Editor to open a word processor, spreadsheet, or text file in HoTMetaL PRO. You can also convert multiple files using Convert Documents... in the Tools menu in HoTMetaL PRO Information Manager. HoTMetaL PRO uses built-in filters to automatically convert each file to HTML as it opens them. Each block of text is converted to the appropriate element: paragraphs are converted to P elements, headings to H1-H6, numbered lists to OL, unnumbered lists to UL, tables and spreadsheet information to HTML tables, etc.

Supported formats

By default, HoTMetaL PRO will convert documents from the following formats:

- plain text (text that is not marked up in HTML)
- AMI Pro (1.1, 2.0, 3.0)
- MS Word for PC (2.0, 3.0, 4.0, 5.0, 5.5)
- MS Word for Windows (2.0, 6.0, 7.0., Word 97)
- RTF
- WordPerfect for DOS/Windows (4.2, 5.0, 5.1, 5.2, 6.0, 6.1, 7.0)
- MS Works for Windows (3.0, 4.0)

If you chose Custom Install when you installed HoTMetaL PRO, any additional conversion formats you selected at that time will also be supported. If you did not install these additional filters and wish to do so, do the following:

- Start the installation program again and choose Custom Install on the Setup Type installation panel.
- On the Select Components panel, choose Other Formats from the components list.

If you do not want to completely reinstall HoTMetaL PRO; deselect all of the other product components.

The following additional formats are available from the custom installation:

Cliq-Word	NBI Net Archive
DECdx (WPS-PLUS)	ODA FOD26
HP Word PC	Office Writer
IBM DCA-RFT	Quadratron Q-ONE
IBM DCF Script	Uniplex II Plus
IBM DCA-FFT	Wang OIS/VS Com
IBM Displaywrite	WordMARC
MASS-11	WordPerfect for Mac
MS Word for Macintosh	WordPerfect 4.2
MultiMate	WordStar
Navy DIF	Wang WITA

If you try to open a file that is not in one of the supported formats, an error message will appear and the file will not open. However, you may be able to open the file in another application and save it in a format that is supported.

Converting documents with graphics

If a document contains a graphic in one of the supported graphics formats, it will be converted to GIF format and saved in a separate file. An IMG element whose URL contains the full path to the GIF will be placed in the converted document. HoTMetaL PRO will convert graphics from the following formats:

- BMP
- GIF
- JPEG (JFIF)
- MacPaint
- PCX
- TIFF

In addition, HoTMetaL PRO can convert the raster, or bitmap-oriented, portions of the following complex graphic types: EPS, PICT, RTF, SDW (AMI Draw), WMF, and WPG (WordPerfect Graphics). HoTMetaL PRO will not convert vector, or outline-oriented, graphics such as embedded pie-charts or equations from these formats.

Graphics that are linked to word processor files (rather than inserted directly into the file) will *not* be converted.

Converting a single document

You can convert a single file simply by opening it in HoTMetaL PRO Editor. The file converts as it opens and then you can make any necessary changes.

To convert a file:

- Choose Open... from the File menu, type Ctrl+O, or click on the toolbar button. The Open dialog box appears.
- Choose the appropriate format from the File Type drop-down list.
- Select the document to be converted and click on Open. The document is automatically converted as it opens.

Converting multiple documents

If you have many files to be converted, you can convert them all at once in the Information Manager. You should open a single file first and make any necessary changes to optimize the conversion (see this page).

To convert multiple files:

- Choose Convert Documents... on the Tools menu.
- Select the documents to be converted and click on Open . The Choose Target Folder dialog box appears.
- Select the folder you wish the converted documents to be placed in, and click on OK .

Preparing files for conversion

Because word processor files tend to be less structured than HTML, you should expect to do some refinement of the markup in the Editor. Also, the conversion doesn't use the full HTML element set and you may wish to use richer markup.

If you have a lot of material to convert, it would be worthwhile to test the conversion process. Convert one or two files and have a look at the results. You may find that manipulating the styles in your word processor first gives better results. For example, if your headings are the same point size as your paragraphs, but bold, the conversion process may have trouble recognizing them. If you set the font sizes for each heading to be much larger that lower level headings, the conversion program will provide better results.

The conversion program uses various guidelines to decide how to handle text. You can configure how HoTMetaL PRO converts paragraphs, blockquotes, and headings.

Setting conversion options

You can configure the file conversion settings used when you open a word processor file in HoTMetaL PRO. In general, HoTMetaL PRO uses length (in characters), font size, and font style to distinguish paragraphs from headings.

By default, HoTMetaL PRO regards paragraphs in an emphasized (bold, italic, or bold-italic) font style as possible headings. Unless you specify otherwise, the *first* emphasized paragraph in the document with a length less than, or equal to, the maximum heading length will be surrounded with the H1 element. Any subsequent paragraphs with the same font and font style as the first H1 will also be converted to H1. This pattern is then followed for the other heading levels.

To set document conversion options:

- Open the HoTMetaL PRO Information Manager.
- Choose Document Conversion Options... from the Tools menu.

The following settings are available:

- Use lists – If you want HoTMetaL PRO to look for list items, turn on this check box.
- Minimum number of list items – Specify the number (between 1 and 4) of one-line paragraphs required to be considered a list.
- Use blockquotes – If you want HoTMetaL PRO to look for blockquotes, turn on this check box. HoTMetaL PRO will surround a paragraph with a BLOCKQUOTE element if its length is greater than or equal to the minimum blockquote length, it is not in an emphasized font style, it ends with a 'return' character but does not contain any other return characters, and it contains start and end quotation marks. The acceptable quotation characters are straight double quotes, opening and closing double quotes, European quotes (double-angled brackets), and smart quotes.
- Minimum blockquote length – HoTMetaL PRO considers only paragraphs that have the specified number of characters (or less) to be blockquotes. This helps avoid ordinary lines of text being interpreted as quotes. The possible values for this setting are 5 to 32767 characters, inclusive.
- Allow headings in tables – If you want HoTMetaL PRO to look for headings in tables, turn on this check box.

- Maximum heading length – HoTMetaL PRO considers only paragraphs that have the specified number of characters (or less) to be headings. This helps avoid ordinary lines of text being interpreted as headings. The possible values for this setting are 5 to 32767 characters, inclusive.
- Heading 1 to Heading 6 font size – Specify the font size that you want HoTMetaL PRO to recognize as each level of heading.
- Bullet characters – The default characters that HoTMetaL PRO recognizes as the start of a bullet item are '-', '*', and '.'. Add any additional characters that have been used as bullets in your document.