

Change History for RTFtoHTML

The latest version of this document and other information about RTFtoHTML is kept at <http://www.sunpack.com/RTF>. You may contact the author: Chris Hector at chris@sunpack.com

Features in 3.9

Added support for >99 split files

The string "ShortFileDigits" sets the number of digits to use in short file names. The default value of 2 supports 99 split files, 3 supports 999 files.

Switch allows empty paragraphs to be output

The string "DoAllPars" if set to 1 will allow empty paragraphs to be output. Normally, these are suppressed (which is a nice feature within lists - otherwise you get empty bullet or numbered list entries.)

**
 is generated only for empty CELLS**

Previous versions generated a trailing
 in all cells to work around a bug in Netscape that caused empty cells to lose their border. Now only empty cells get the BR so that no empty space is generated at the end.

Changed *-sym mapping for 0xcd to reflexsubset

Bugfixes in 3.9

Cells are aligned at the top

The <TD> tag now has the valign=TOP attribute. This matches how MSWord aligns cell contents.

**Bullets were being preceded by extra
**

This fails when in mid sentence - .removed it.

Fixed improper restarting of outer list levels

This showed up as a list item without a in front following a nested list.

Corrected misuse of Charset support for Macintosh.

The 3.8 feature of allowing multiple character sets worked with Macintosh documents containing Windows (ANSI) characters - the other way around now works.

Changed first footnote to put it within the enclosing <P>

The first footnote (and sometimes the first paragraph) of a document could have text that preceded the <P> starting tag.

Features in 3.8

Footnote separator added

The string in html-trn "FNSEP" can be used to specify the text that is inserted between the text of the document and the footnotes.

In Frames generation, external links have a TARGET attribute

"The string in html-trn "ExtTarget" can be used to specify the text of a TARGET option. This allows external links to (links pointing outside of the current document) to be loaded in a full window rather than the old default behavior of loading the reference into the current frame.

added Switch in html-trn (PreferrEmbed) to select embedded graphics over external links

In an RTF document, you can link to an external graphics file. When these links exist, the RTF file may contain a copy of the graphic file internally. The default behavior of RTFtoHTML is to ignore the embedded graphic and generate a link to the graphic file. (This can save space if you re-use the graphic multiple times.) You can now define the string "PreferrEmbed" to 1 and the filter will choose the internal RTF copy of the graphic.

Added support for Red text that was not exactly red

Some platforms generate a color table where 'red' text has an RGB representation which is close to "FF0000", but not exact. The filter would not recognize this text as 'red' and therefore not treat it as a cross-reference to a URL or heading. Now the filter finds the closest match to red in the color table and treats it as red.

Bookmarks are now translated to anchors

A bookmark in RTF now generates an anchor tag <A>bookmarkname at the start of the bookmarked text.

Symbols in Word 7 documents (created with INSERT SYMBOL) are now supported.

Word version 7 used field codes to represent a symbol character. The implementation was incorrect leaving no 'result' value in the RTF. Now the filter recognizes this error, and generates the proper field result.

The -s option now forces Macintosh PICT format files to have an extension of '.PIC' instead of '.PICT'

Heading support now extended to allow style names with aliases.

Word allows a style name to be specified as "heading 6,COM 1H,C.head1" which means that the style name is "heading 6" and it can also be referenced by the aliases "COM 1H" and "C.head1". These 'alias' style names are now supported by having the .Pmatch table specify any of the allowable aliases.

Character translation of NON-ISO characters (symbol font) can now generate UNICODE representation.

Thanks to Alan Flavell the output translation and conversion has been modified to produce output that conforms to RFC2070/HTML4.0 standards, using, where necessary, unicode numeric character references (&#bignumber; representations). The html-map.v4 file is the UNICODE version of the mappings. These mappings require support from browsers - the default version will work on older browsers (as well as the new ones.) The filter ships with a "practical" version of this translation, a more accurate is the "pedantic" version available at <http://www.physics.gla.ac.uk/r2h-extras/>. The "practical" variants of these files contain various dumbing-down from the "pedantic" versions, in the interest of getting something to be actually displayed on some current browsers (e.g a normal asterisk in place of the mathasterisk, etc.).

Translation files can now be placed in the same directory or folder as the input RTF file.

This allows different translation options to be used for different RTF input. (For example different authors/conventions or output styles.) The filter first looks in the directory containing the input file, then in the current directory (UNIX) then the directory containing the filter itself.

Multiple Character set support

If an RTF file was created on a Macintosh and then moved and edited on a PC (as well as the reverse scenario), the RTF file could contain both character sets. (Macintosh and PC character encodings are different for some characters above decimal 128. RTFtoHTML now recognizes and correctly interprets both encodings and switches from one to the other correctly.

Bugfixes in 3.8

An anchor (font _Name) following a list caused an extra, unwanted .

Headings following a list caused the closing tag markup not to be correctly nested.

Footnotes now handled after an italic

A bug in footnote processing caused some footnotes to be ignored when following styled text (bold/italic/fon-changes)

Anchors in tables now handled properly

Index entries in or preceding footnotes caused a spurious "]" to be generated.

Footnotes in Headings no longer cause an abort

red text as second character after a heading, caused the first character to be discarded

The _Literal markup was generating <P> marks at end of line instead of a new-line character.

_Literal is a way to allow the user to specify literal HTML markup. The filter adding <P> markup could result in incorrect HTML.

Index entries within headings were causing the heading to be split at the point of the index entry.

Core dumps caused from table processing are fixed.

Red text that is not a URL, Email address or heading references is now ignored (after a warning message.)

Fixed bug in tables not terminating text markup

Fixed termination of unix/PC versions on warnings

Warnings on buffer overruns, (caused by incorrect RTF markup, or html-trn settings) now include the offending text. They also do some cleanup to allow translation to continue (although some of the offending text may be discarded.)

Footnote text markup not flushed at EOF leaving unclosed tags.

Text markup at EOF was not closed when followed by footnotes.

double underline at end of paragraph genereates an orphan ''

_Name in table not properly terminated

Footnote in table cell not terminated correctly

What happened to 3.7?

The 3.7 release was never made public. All changes between 3.6 and 3.8 are listed above.

Bugfixes in 3.6

Footnotes were lost in the 3.5 version

A regression in version 3.5 caused footnotes to be lost. The footnote references appeared, but not the underlying text.

Field Contents were lost in the 3.5 version

Fields containing non-graphic results were being lost in version 3.5. In earlier versions (3.0.1) fields could cause the remainder of the document to be discarded. Now fields are properly parsed.

Linked Pictures could result in double images

When an image was linked into a document AND the image itself was included, the filter would generate a link for each of these. Now the filter ignores the link to the image and uses the imbedded version of the image.

Hot text, footnotes or names prior to Headings corrected

When hot text, footnotes or Names prior to a heading was encountered, the heading was incorrectly imbedded in the previous markup.

Bugfixes in 3.5

Accent character was being dropped from Headings/Title

The accent character ' or right quote was being dropped from Headings and Title. There were several other 'special' characters that were also dropped. These now appear correctly.

The navigational text (Contents, Index...) were not correct on the second RTF file translated

When RTFtoHTML was invoked to translate more than one document, the links/titles for table of contents and index appeared blank after the first document. This was caused by a flag not being re-initialized correctly.

Empty heading paragraphs resulted in invalid HTML , crashes or (post 3.4) 'mismatch' errors.

An internal error was causing the first pass to process empty heading paragraphs differently than the second pass. Prior to 3.4 this could result in crashes. In 3.4 the problem was detected and the filter would issue an internal error and exit. The mismatch is now corrected.

Added "Body Text"

The style "Body Text" was added to html-trn and mapped to "Normal".

Using the -h option could result in lost sections and invalid links.

An error occurred if two headings in a document contained the same leading text, but one had lower case characters and the other had upper case. The file name processing assumed that the filename 'SomeHead.rtf' was different than 'somehead.rtf'. On Macintosh, DOS and Windows systems this is not true, so the filter would overwrite one file with another. The filter now makes these names unique (even on Unix platforms.)

Version 3.4 could fail on locked (or full) disks.

In 3.4 the library tmpfile() was used to create temporary files. However in some installations, this would attempt to create temp files on a locked or full disk. Now temp files are created in the same directory as the output files (which will be writeable.)

Version 3.5 could abort on empty tables.

Some tables (empty tables, or tables that were exactly the same as a prior table) could cause aborts of the filter.

Features in 3.4

Added Win32 (native Win95 and WinNT) binaries

The Windows 95 and Windows NT binaries are console applications, which mean that they can still be run from a DOS prompt. They are 32-bit applications which should improve performance. They also are distributed with batch files which support drag-n-drop conversions, and allow you to specify default conversion options.

Added Long Filenames to Win32 versions

The Windows 95 and Windows NT versions of the filter now support long filenames. By default, the long filenames are used, but short filenames can still be forced by the -s command line option.

Error messages captured to a .err file

All of the error and warning messages are emitted to standard error as before, but now are also saved to a .err file. This is particularly useful for DOS/Win/Win95 installations where stderr cannot be re-directed to a file. It is also helpful for converting several document - since each document gets its own .err file.

Added support for Office 97 (RTF 1.5)

In version 3.0.1, converting a document produced by Office '97 would generate errors such as:

```
ReadStyleSheet: unknown token "\adjustright"  
ReadStyleSheet: unknown token "\cgrid"
```

In some cases the conversion would still be successful, but in others, part of the RTF input would be ignored.

Office '97 documents are now completely processed, with no warning messages.

Heading numbering is now preserved.

An often requested feature was to preserve the paragraph numbering on headings (heading 1, ... heading 6 styles). In previous versions, the filter stripped these heading numbers. Now the heading numbers appear in the body of the document as well as the table of contents. Note that if you use the feature of linking to headings with red text, the linking text must now contain the heading numbers as well.

IMG HEIGHT/WIDTH tag improvements

The HEIGHT and WIDTH tags generated for embedded pictures (markup) have several changes:

HEIGHT and WIDTH tags can be disabled

By setting "SkipDimsOnIMGs" to 1 in the strings table, you can force the filter to not generate HEIGHT and WIDTH tags. This feature depends on the new ".Strings" section of the html-trn file. Sample use is as follows:

```
.Strings  
"SkipDimsOnIMGs",1
```

HEIGHT and WIDTH tags now use scaling information

If an imbedded image was enlarged or reduced within the RTF document, the HEIGHT and WIDTH tags now are generated using the scaled sizes. This information will allow the browser to scale the image when it is displayed.

WMF sizing improved

HEIGHT and WIDTH tags for WMF files were incorrectly generated (although this scaling matches how they are treated by Word 6 on the Macintosh.) There is now an additional scaling factor (133%) applied to the HEIGHT and WIDTH tags of WMF files. This scaling factor can also be changed my modifying the following line in html-trn:

```
"WMFAdjust",133
```

The 133 indicates that WMF HEIGHT and WIDTH tags should be 133% larger than would be generated normally by the filter. This value appears to be correct (from the testing that I have done) and should not have to be modified.

Navigation Panel generation can be disabled

By default, all translated documents that are split into multiple documents by the -h option (Split Level) have a navigation panel to get from one document to another. This can be disabled by setting "SkipNavPanel " to 1 in the .Strings table of html-trn:

```
.Strings  
"SkipNavPanel",1
```

Leading table of contents generation can be disabled

By default, if any headings are found in a document, there will be an internal table of contents generated at the start of the HTML document. This can be disabled by setting "SkipLeadingToc" to 1 in the .Strings table of html-trn:

```
.Strings  
" SkipLeadingToc ",1
```

Trailing table of contents generation can be disabled

By default, if any headings are found in a document, there will be an internal table of contents generated at the end of the HTML document. This can be disabled by setting "SkipTrailingToc" to 1 in the .Strings table of html-trn:

```
.Strings  
" SkipTrailingToc ",1
```

Dropped "- Title" from document titles

In 3.0.1, the filter generated "- Title" at the end of document titles for the title page. This was designed for split documents, but also appeared in non-split documents. Now the title for the Title page is simply the title from the RTF document (or the command line using the -T option.)

Default Title is now null

In 3.0.1 if no title was supplied in the RTF document or on the command line (with the -T option), the filter supplied "No Title" for a default. Now the default title is empty. This allows browsers to substitute the filename for a title. Alternatively, you can set a default title in the .Strings table of html-trn:

```
.Strings
" DefaultTitle ", "My Default Title"
```

Tables can be forced to have borders.

By default, tables generated by the filter will have borders turned on if the first row of the table contained a border. If you want to override this checking, you can force all tables to be generated with borders by setting "AllBorder" in the .Strings table of html-trn:

```
.Strings
" AllBorder ", 1
```

Improved error processing in macintosh versions

In 3.0.1 if multiple files were dropped on the filter, a fatal error in an document would not correctly terminate all subsequent processing. This could result in invalid conversions, application aborts or freezes. Now a fatal error in a document will terminate processing. You will need to drag-n-drop the remaining files on to the filter.

Footnotes not starting at 1 are supported

In version 3.4 and above, automatically numbered footnotes that start at a number other than the default(1) are now supported. Note that footnotes that are reset on page boundaries or section boundaries (in the RTF) will continue to increment throughout the document in the HTML version - since there are no page-breaks in HTML.

Bugfixes in 3.4

Headings following tables now generate correct markup

In 3.0.1 a table that was immediately followed by a heading would not generate correct markup. The heading would be incorrectly included in the table.

Win/DOS 8 character filename bug fixed.

In Windows and DOS versions, footnotes were accumulated in a file whose name could be too long if the original input file was 8 characters (+3 character extension.) This could result in a variety of conversion time errors. Now a correct temp file name is used.

Headings in tables problem corrected

In 3.0.1, if a heading appeared within a table, the table markup would be incorrectly generated. The table itself could be split (incorrectly) across multiple files. Now, headings found within tables will not be used for file splitting, and will not appear in the table of contents.

Font Sizes in Tmatch table are corrected

In 3.0.1 the font size in the Tmatch table needed to be twice the point size in order to match, thus a 10 point font would require a 20 point Tmatch entry. Now the point size matches correctly.

Features in 3.0.1

- Σ Added HEIGHT and WIDTH tags to regular inline images. This will improve the performance for most browsers, since they do not need to download the images prior to rendering the entire page. It also improves the rendering, since images that were re-sized in the RTF document, will now be displayed in HTML at the correct size. Note that cropped images may not be properly cropped, since that can only be done by the graphics filtering program. Also, images that are not imbedded in the RTF file (linked graphics) do not generate HEIGHT and WIDTH tags.
- Σ Added the ability to process multiple files on UNIX and under DOS. This is done by adding additional options/filenames on the end of the command.
- Σ RTF files containing links to images will now have (inline images) generated instead of <A HREF...> (links).
- Σ The -h0 (split level) option is now automatically set if any of the following are specified. If a different split level is specified by the user, it takes precedence.
- Σ -c (Table of Contents),
- t (Refs on top)
- F (Frames)
- N (Custom Navigation Panel)
- x (Index)
- Σ Made -F (Frames) imply -c (Table of Contents)
- Σ Changed numeric suffixes for output file names so that they sort in numerical order.

Bugs Fixed in 3.0.1

- Σ Fixed a bug in Mac version where -G was not being added
- Σ If the output filename is specified with a .htm extension, that extension is now used for all HTML filenames.
- Σ Fixed unix makefile to build rfttohtml.bin and shell to invoke it
- Σ modified htmlunix.c and fileops.c to allow passing of LIBDIR as non-string
- Σ Added bugfix to correct NULL font name problem (ewiles@mclean.sterling.com)
- Σ Corrected a bug in footnotes which caused footnote reference numbers to be off by one. It was a increment variable side effect that caused the problem, so it only appeared in some compilers!
- Σ Fixed a bug in XformStr where global substitutions were not correctly handled (move to end of string was bad value). This showed up when links to pictures were used.
- Σ Corrected a bug that caused visible index entries to be discarded (even though the index was correctly generated.)
- Σ fixed regexp prototypes in Unix source distribution
- Σ Added nav-panl and graphic files to Unix distributions
- Σ changed standardcharnames.h to stdchar.h (in d and e)
- Σ Default GIF files used in the navigation panel now have 8 character (or less) file names to allow them to be used and properly distributed on DOS/WIN 3.1 platforms.
- Σ Modified the file names to eight characters for translation files:
html-trans becomes html-trn and nav-panel becomes nav-panl

Features in 3.0c

- Σ The filter now supports RTF version 1.4 (Microsoft Word version 7) files.

Bugs Fixed in 3.0c

- Σ Lists are now supported within table cells.
- Σ Mac 68k version had intermittent memory errors resulting in freezes and system crashes. This has been corrected.
- Σ Mac version progress bar was not displayed until the second pass through the file. This was only noticeable on very large files, but now both passes are represented.

Version 3.0b

- Σ Modified table support to allow images
- Σ Modified timing to improve performance on the Macintosh and give more cycles to the filter when running in the background.

Changes in RTFtoHTML 3.0a

Features

- Σ Added support for HTML 3.0 tables
- Σ RTFtoWEB support incorporated into the main filter. All platforms now have identical functionality. The RTFtoWEB code was written by Christian Bolik, and it includes many features:
 - Σ A single RTF document can be split into a collection of HTML documents that are linked together by a navigation panels, and a common Table of Contents and Index.
 - Σ The Table of Contents and Index are generated automatically and contain hypertext links to the appropriate sections within the document.
 - Σ fully customizable navigation panels on top and bottom of each page
 - Σ active Cross references to headings, mail addresses and URLs
 - Σ support for a few of Netscape's HTML extension, such as the `<center>` tag and background images
 - Σ Netscape 2.0 Frames are supported to allow the Table of Contents to appear as a separate frame.
 - Σ Short file names can be generated for DOS and Win 3.1 environments.
 - Σ PowerPC native support for Macintosh
 - Σ Added support for subscript and superscripts
 - Σ Error messages now display text prior to where the error occurred so that you can find the location of the problem in your input source.
 - Σ Changed footnotes processing so that a separate file is created only when file splitting is enabled. Footnotes are now always imbedded in the current file at the end. They also have a "back" feature that sends you back to the origination when you click on the footnote.
 - Σ Documents that link to graphic files (instead of imbedded graphics) are now supported.
 - Σ Character translation has been improved by providing a complete mapping for ansi and Macintosh character sets.

Bugfixes

- Σ Corrected bug that caused text to be truncated on 80 column boundaries.
- Σ Corrected bug that caused hard-return (\line) to be lost
- Σ Removed the error which caused the following messages :
 - Σ ReadStyleSheet: unknown token "\widctIpar"
 - Σ Unknown symbol... near line .. position
- Σ The filter now correctly dealt with documents that contain revision marks.
- Σ Footnotes where the footnote reference was formatted differently in the footnote, than it was in the body of the document - are now processed correctly.
- Σ Corrected many memory overruns that caused errors and crashes.
- Σ Long lines of text (>80)with no spaces are now properly handled.
- Σ Table of contents now generates properly nested anchors in all cases.