NAME

unroff-html-ms – back-end to translate 'ms' documents to HTML 2.0

SYNOPSIS

unroff [**–fhtml**] [**–ms**] [*file* | *option...*]

OVERVIEW

When called with the **–fhtml** and **–ms** options, the troff translator *unroff* loads the back-end for converting "ms" documents to the Hypertext Markup Language (HTML) version 2.0.

Please read **unroff**(1) first for an overview of the Scheme-based, programmable troff translator and for a description of the generic options that exist in addition to **–f** and **–m**. The translation of basic troff requests, special characters, escape sequences, etc. as well as the HTML-specific options are described in **unroff-html**(1). For information about extending and programming *unroff* also refer to the *Unroff Programmer's Manual*.

OPTIONS

The **–ms** extension provides a number of keyword/value options in addition to those listed in **unroff**(1) and **unroff-html**(1):

signature (string)

If non-empty, the value of this option together with a <hr> tag is appended to each HTML output file created. The *substitute* Scheme primitive (as described in the Programmer's Manual) is applied to the value of the option, so that date, time, environment variables, etc. can be interpolated.

split (integer)

This option specifies whether to split the output document into individual files for each major section. If a positive integer *level* is assigned to the option, a new output file is opened for each numbered header (**.NH** request) with a level equal to or numerically less than *level*. Use of this feature requires that the **document** option has bee set, as otherwise the HTML document is sent to standard output. The default value is 0, i. e. all sections will be written to a single file.

toc (boolean)

If true, a table of contents with a hypertext link for each section is generated automatically and inserted after the front matter (title, author information, abstract) and before the first section. Use of this feature requires a non-zero value for the **split** option. The default is to produce a table of contents if **split** is non-zero.

toc-header (string)

This option defines the contents of the <h2> header element prepended to an automatically generated table of contents. Its value is subject to a call to *substitute*. The default is the string "Table of Contents".

pp-indent (integer)

The number of non-breakable spaces (as specified by the predefined Scheme variable *nbsp*) to generate for a paragraph created by the **.PP** macro. The default is 3. This option, as well as **signature**, is typically set in the user-preferences file **~/.unroff**, or in a document-specific Scheme file or at the beginning of the document proper.

footnotes-header (string)

The contents of the <h2> header element prepended to the footnotes section that is appended to the document if any footnotes were used, and that also appears in the automatically generated table of contents. As with all string option listed in this section, the *substitute* primitive is applied to the option's value. The default is the string "Footnotes".

footnote-reference (string)

This option controls the text generated by each use of the variable '**', which produces a footnote (hypertext) reference. Its value is passed to a call to *substitute* with the current footnote number as another argument, so that the specifier "%1%" can be used to interpolate the footnote number. The default is the string "[note %1%]".

footnote-anchor (string)

This options specifies the footnote reference that appears at the beginning of each footnote proper if **.FS** was called without an argument. The option's value is passed to a call to *substitute* with the footnote number generated by the last use of '**' as another argument. The default is "[%1%]".

FILES

unroff reads and parses an "ms" document composed of one ore more input files. As usual, the special file name '**–**' can be used to interpolate standard input. If no file name is given in the command line, *unroff* reads from standard input.

The resulting HTML document is sent to standard output, unless a file name prefix is assigned to the **document** option. In the latter case, the **split** option controls splitting of the output into separate files at section boundaries as described under OPTIONS above. A number of other features, such as footnotes, also require that the **document** option is supplied, as separate output files are created for them (regardless of the value of **split**). In any case, the name of each output file consists of the value of **document**, followed by an optional suffix, followed by the extension ".html".

EXAMPLE

To translate an "ms" document composed of several input files, *unroff* could be invoked like this:

unroff –fhtml –ms document=thesis split=2 intro.ms 1.ms 2.ms 3.ms app.ms

The names of all output files will have the prefix "thesis", and the resulting HTML document will be split into separate files at each level 1 section or level 2 section.

DESCRIPTION

The following **–ms** macros are translated (in addition to any user-defined macros):

.AB	.AE	.AI	.AU	.B	.B1	.B2
.BD	.BX	.CD	.DE	.DS	.FA	.FE
.FS	.I	.ID	.IP	.LD	.LG	.LP
.NH	.PP	.PX	.QP	.R	.RE	.RS
.RT	.SH	.SM	.TL	.UL	.UX	.XA
.XE	.XS					

These predefined strings and number registers are recognized:

*-	*(DY	*(MO	*Q	*U	\n(PN

In addition, a number of macros are either silently ignored or cause a warning to be printed, because their function either cannot be mapped to HTML 2.0 elements or assumes a page structure:

.AM	.BT	.CM	.CT	.DA	.EF	.EH
.HD	.KE	.KF	.KS	.ND	.NL	.OF
.OH	.P1	.PT	.TM	.MC	.1C	.2C

The font switching macros are based on changes to the fonts 'R', 'I', and 'B', as explained under FONTS in **unroff-html**(1). Of course, this fails if the fonts (which are mounted on startup) are unmounted by explicit **.fp** requests.

Upper or lower case letters are accepted as section numbers by **.NH** when the argument "S" is used to set new section numbers. This is useful for appendices and similar constructs.

The translation rule for **.IP** employs a heuristic to determine whether to generate a definition list or an unordered list: if the first in a sequence of indented paragraph macros is called with a tag consisting of one of the special character \(bu or \(sq, a definition list is begun, otherwise an unordered list. Since exdented[*sic*] paragraphs cannot be expressed in HTML 2.0, a warning message is printed when a call to the macro **.XP** is encountered.

All footnotes are concatenated and placed in a separate output file, and a corresponding section (with a user-defined header) holding the footnotes is appended to the document automatically. Use of the string '**' generates a hypertext link to the beginning of the footnote created by the next call to **.FS** and **.FE**. The actual text generated by using '**' as well as the footnote reference that appears in the footnote proper

are controlled by two options as explained under OPTIONS above. A warning message is printed on termination if '**' has been used but a corresponding footnote was not seen. As an alternative to '**', the new request **.FA** can be used to produce a footnote anchor together with a hypertext link; the anchor is the argument to the macro (however, '**' itself must not be used in a call to **.FA**).

Likewise, a hypertext reference is created for each use of the table of contents macros **.XS** and **.XE** (optionally accompanied by calls to **.XA**).

SEE ALSO

unroff(1), **unroff-html**(1), **troff**(1), **ms**(5 or 7).

Unroff Programmer's Manual.

http://www.informatik.uni-bremen.de/~net/unroff

Berners-Lee, Connolly, et al., HyperText Markup Language Specification—2.0, Internet Draft, Internet Engineering Task Force.

BUGS

The macro **.UL** is currently mapped to a call to **.I**, as underlining is not supported by the HTML back-end of *unroff* 1.0.

Footnote references and requests such as **.sp** that cause non-character-level markup to be generated must not be used inside a numbered header.

When creating a hypertext anchor for **.XS** and **.XE**, there is nothing to put inside the <a> element; therefore a non-breaking space is used.

Changing the number register format of 'NH' to get roman or alphabetic section numbers does not work, obviously.