NAME

man2html – convert UNIX nroff(1) manual pages to HTML format

SYNOPSIS

man2html [**-aliases** *file*] [**-bare**] [**-belem** *name*] [**-botm** *lines*] [**-cgiurl** *string*] [**-cgiurlexp** *expr*] [**-compress**] [**-externs** *file*] [**-headmap** *mapfile*] [**-help**] [**-index**] [**-k**] [**-leftm** *chars*] [**-leftside** *file*] [**-nodepage**] [**-noheads**] [**-pgsize** *lines*] [**-seealso**] [**-solaris**] [**-sun**] [**-title** *string*] [**-toc**] [**-topm** *lines*] [**-uelem** *name*]

Typical Usage:

man2html [**-options**] **<** *infile* **>** *outfile*

man *topic* **|** **man2html** [**-options**] **>** *outfile*

DESCRIPTION

The **man2html** filter reads formatted nroff text from standard input (*stdin*) and writes a HTML document to standard output (*stdout*).

The formatted nroff output is surrounded with **<PRE>** tags with the following exceptions/additions:

- Section heads are wrapped in HTML *header* tags. See the **SECTION HEAD MAP FILE** section below for additional information. The **-noheads** option can be used to disable this feature.
- Bold words designated by a "<char><bs><char>" sequences are wrapped in **<B>** tags (or the element specified via the **-belem** option).
- Underlined words designated by a "_<bs><char>" sequences are wrapped in **<I>** tags (or the element specified via the **-uelem** option).

OPTIONS

-aliases *file*
: Gives man2html a list of manpage aliases, i.e., alternate names for manual pages. The file contains one alias per line, followed by the actual manual page name. Man2html will link the reference for the alias to the actual manual page.

-bare
: This option will eliminate HTML **<HEAD>** and **<BODY>** tags from the output. This is useful when you wish to incorporate the output into another HTML document.

-belem *name*
: Use *name* as the name of the element to wrap overstriken characters. The default is **B**.

-botm *lines*
: The *lines* argument specifies the number of lines representing the bottom margin of the formatted nroff input. The line count includes any running footers. The default value is 7.

-cgiurl *string*
: The *string* argument specifies a template URL for creating links to other manpages. See the **LINKING TO OTHER MANPAGES** section below for additional information.

-cgiurlexp *expr*
: The *expr* argument specifies a Perl expression evaluting to a URL for creating links to other manpages. See the **LINKING TO OTHER MANPAGES** section below for additional information.

-compress
: Compress consecutive blank lines into a single line.

-externs *file*
: Tell man2html to not attempt to link to the manual pages listed in the file.

-headmap *mapfile*
The *mapfile* argument is read to determine which HTML header tags are to be used for various section heading in the manpage. See the **SECTION HEAD MAP FILE** section below for information on the format of the map file.

-help Print out a short usage message and then exit immediately.

-index Write an un-numbered list of headers at the end of the output HTML, which can be used as-is, or post-processed to make that list an HTML "nav" area that can be manipulated with CSS.

-k Process input resulting from a manpage keyword search (**man -k**). See the **KEYWORD SEARCH** section below for additional information.

-leftm *chars*
The *chars* argument specifies the width of the number of characters making up the left margin of the formatted nroff input. The default value is 0.

-leftside *file*
Man2html detects references to other manual pages by manpage name (a word, possibly with embedded "–") in bold font, followed by a section number in parentheses. If the manpage name is split by nroff's hyphenation, man2html can be confused, because it cannot tell whether a "–" at the end of the line is part of a manpage name, or just the hyphenation marker.

The **-leftside** option allows you to tell man2html about these special cases, using a file containing the left-side of hyphenated manpage names. If the option is not given, man2html uses a short fallback: "apt", "cvs", "git", "sccs' and "sysv".

-nodepage
By default, **man2html** merges multi-page formatted nroff into a single page. This option may be used to disable depagination, causing running headers and footers in the formatted nroff input to be carried over into the HTML output.

-noheads
By default, **man2html** wraps section heads in HTML header tags. See the **SECTION HEAD MAP FILE** section below for additional information. This option may be specified to disabled this feature.

-pgsize *lines*
The *lines* argument specifies the number of lines making up the page size (length) of the formatted nroff input. The default value is 66.

-seealso
If the **-cgiurl** option has been specified, then this option restricts the creation of links to other manual pages to the **SEE ALSO** section only.

-solaris
If the **-k** option has been specified, then this option modifies its operation to process the alternate manual page keyword search format produced by the **man**(1) utility on systems running *Solaris*. See the **KEYWORD SEARCH** section below for additional information.

-sun Do not require a section head to have bold overstriking in the formatted nroff input. The option is called **sun** because it was on a Sun workstation that section heads in manpages were found to not be overstruck.

-title *string*
By default, **man2html** does not generate a HTML title (**<TITLE>**). This option sets the title of the HTML output to the specified *string*.

-toc Generate IDs compatible with HTML::Toc, which is used to produce a separate table of contents file. The IDs embedded in HTML header-tags, e.g., "<h3>" will have a "–toc" suffix.

-topm *lines*
The *lines* argument specifies number number of lines representing the top margin of the formatted nroff input. The line count includes any running headers. The default value is 7.

-uelem *name*
Use *name* as the name of the element to wrap underscored characters. The default is **I**.

SECTION HEAD MAP FILE

The **-headmap** option may be used to customize which HTML header tags, **<H1> ... <H6>**, are used in manpage section headings. Normally, **man2html** treats lines that are flush to the left margin (**-leftm**), and contain overstriking (overstrike check is canceled with the **-sun** option), as section heads. However, you can augment/override what HTML header tags are used for any given section head.

In order to write a section head map file, you will need to know about **perl**(1) associative arrays. You do not need to be an expert in **perl** to write a map file, however, having knowledge of **perl** allows you to be more clever.

Augmenting the Default Map

To add to the default mapping defined by **man2html**, your map file will contain lines with the following syntax:

```
$SectionHead{'<section head text>'} = '<html header tag>';
```

where

<section head text>
is the text of the manpage section head. For example: **SYNOPSIS** or **DESCRIPTION**.

<html header tag>
is the HTML header tag to wrap the section head in. Legal values are: **<H1>**, **<H2>**, **<H3>**, **<H4>**, **<H5>**, **<H6>**.

Overriding the Default Map

To override the default mapping with your own, then your map file will have the following syntax:

```
%SectionHead = (
    '<section head text>', '<html header tag>',
    '<section head text>', '<html header tag>',
    # ... More section head/tag pairs
    '<section head text>', '<html header tag>',
);
```

The Default Map

As of this writing, this is the default map used by **man2html**:

```
%SectionHead = (
    '\S.*OPTIONS.*'             => '<H2>',
    'AUTHORS?'                  => '<H2>',
    'BUGS'                      => '<H2>',
    'COMPATIBILITY'             => '<H2>',
    'DEPENDENCIES'              => '<H2>',
    'DESCRIPTION'               => '<H2>',
    'DIAGNOSTICS'               => '<H2>',
    'ENVIRONMENT'               => '<H2>',
    'ERRORS'                    => '<H2>',
    'EXAMPLES'                  => '<H2>',
    'EXTERNAL INFLUENCES'       => '<H2>',
    'FILES'                     => '<H2>',
    'LIMITATIONS'               => '<H2>',
    'NAME'                      => '<H2>',
    'NOTES?'                    => '<H2>',
    'OPTIONS'                   => '<H2>',
    'REFERENCES'                => '<H2>',
    'RETURN VALUE'              => '<H2>',
    'SECTION.*:'                => '<H2>',
    'SEE ALSO'                  => '<H2>',
    'STANDARDS CONFORMANCE'     => '<H2>',
    'STYLE CONVENTION'          => '<H2>',
    'SYNOPSIS'                  => '<H2>',
    'SYNTAX'                    => '<H2>',
    'WARNINGS'                  => '<H2>',
    '\s+Section.*:'             => '<H3>',
);
$HeadFallback = '<H2>';  # Fallback tag if above is not found.
```

Check the **perl** source code of **man2html** for the latest default mapping.

You can reassign the **$HeadFallback** variable to a different value if you choose. This value is used as the header tag of a section head if no matches are found in the **%SectionHead** map.

Using Regular Expressions in the Map File

You may have noticed unusual characters in the default map file, such as "\s" or "*". The **man2html** utility actual treats the **<section head text>** as a **perl** regular expression. If you are comfortable with **perl** regular expressions, then you have their full power to use in your map file.

Caution: The **man2html** utility already anchors the regular expression to the beginning of the line with left margin spacing specified by the **-leftm** option. Therefore, do not use the "^" character to anchor your regular expression to the beginning. However, you may end your expression with a "**$**" to anchor it to the end of the line.

Since the **<section head text>** is actually a regular expression, you will have to be careful of special characters if you want them to be treated literally. Any of the characters '**[ ] ( ) . ^ { } $ * ? + |**' should be escaped by prefixing them by the "****" character if you want **perl** to treat them "as is".

Caution: One should use single quotes instead of double quotes to delimit **<section head text>**. This will preserve any "****" characters for character escaping or when the "****" is used for special **perl** character matching sequences (e.g., **\s**, **\w**, **\S**).

Other Tid-bits on the Map File

Comments can be inserted in the map file by using the '**#**' character. Anything after, and including, the '**#**' character is ignored, up to the end of line.

You might be thinking that the above is quite-a-bit-of-stuff just for doing manpage section heads. However, you will be surprised how much better the HTML output looks with header tags, even though, everything

else is in a **<PRE>** tag.

LINKING TO OTHER MANPAGES

The **man2html** utility allows the ability to link to other manpage references. If the **-cgiurl** option is specified, **man2html** will create anchors that link to other manpages.

The URL entered with the **-cgiurl** option is actually a template that determines the actual URL used to link to other manpages. The following variables are defined during run time that may be used in the template string:

$title The title of the manual page referenced.

$section
The section number of the manual page referenced.

$subsection
The subsection of the manual page referenced.

Any other text in the template is preserved "as is".

Caution: The **man2html** utility evaluates the template string as a **perl** string expression. Therefore, one might need to surround the variable names with '**{ }**' (e.g., **${ title }**) so that **man2html** properly recognizes the variable.

Note: If a CGI program calling **man2html** is actually a shell script or a **perl** program, make sure to properly escape the '**$**' character in the URL template to avoid variable interpolation by the CGI program.

Normally, the URL calls a CGI program (hence the option name), but the URL can easily link to statically converted documents.

Example1:

The following template string is specified to call a CGI program to retrieve the appropriate manpage linked to:

/cgi-bin/man.cgi?section=${section}${subsection}&topic=${title}

If the **ls**(1) manpage is referenced in the **SEE ALSO** section, the above template will translate to the following URL:

/cgi-bin/man.cgi?section=1&topic=ls

The actual HTML markup will look like the following:

<A HREF="/cgi-bin/man.cgi?section=1&topic=ls">ls(1)</A>

Example2:

The following template string is specified to retrieve pre-converted manpages:

http://foo.org/man$section/$title.$section$subsection.html

If the **mount**(1M) manpage is referenced, the above template will translate to the following URL:

http://foo.org/man1/mount.1M.html

The actual HTML markup will look like the following:

<A HREF="http://foo.org/man1/mount.1M.html">mount(1M)</A>

-cgiurlexp

The option **-cgiurlexp** is a more general form of the **-cgiurl** option. **-cgiurlexp** allows one to specify a general Perl expression. For example:

$title=~/^db_/i?"$title.html":"/cgi-bin/man?$title+$section"

A **-cgiurl** *string* can be expressed as follows with **-cgiurlexp**:

return "*string***"**

KEYWORD SEARCH

The **man2html** utility has the ability to process keyword search output generated by the **man -k** or **apropos** commands, through the use of the **-k** option. The **man2html** utility will generate an HTML document of the keyword search input having the following format:

- All manpage references are listed by section.
- Within each section listing, the manpage references are sorted alphabetically (case-sensitive) in a **<DL>** tag. The manpage references are listed in the **<DT>** section, and the summary text is listed in the **<DD>** section.
- Each manpage reference listed is a hyperlink to the actual manpage as specified by the **-cgiurl** option.

This ability to process keyword searches gives nice added functionality to a WWW forms interface to **man**(1). Even if you have statically converted manpages to HTML via another man->HTML program, you can use **man2html** and "**man -k**" to provide keyword search capabilities easily for your HTML manpages.

Processing Keyword Search Results

Unfortunately, there is no standard controlling the format of keyword search results. The **man2html** utility tries it best to handle all the variations. However, the keyword search results generated by the *Solaris* operating system are different enough from other systems that a special command-line option (**-solaris**) must be specified to handle its output.

Example of raw Solaris-type keyword search results:

```
strcpy        strcpy (9f)  - copy a string from one location to another.
strcpy        string (3c)  - string operations
strncpy       strcpy (9f)  - copy a string from one location to another.
strncpy       string (3c)  - string operations
```

If keyword search results on your systems appear in the following format:

<topic> <actual_manpage> (#) - Description

then you need to specify the **-solaris** option in addition to the **-k** option.

ADDITIONAL NOTES

Different systems format manpages differently. Here is a list of recommended command-line options for certain systems:

Convex:	<defaults should be okay>
HP:	**-leftm 1 -topm 8**
Sun:	**-sun** (and **-solaris** when using **-k**)

Some line spacing is lost in the formatted nroff since the spacing would occur in the middle of a page break. This can cause text to be merged that should not be merged when **man2html** depaginates the text. To avoid this problem, **man2html** keeps track of the margin indent right before and after a page break. If the margin width of the line after the page break is less than the line before the page break, then **man2html** inserts a blank line in the HTML output.

A manpage cross-reference is detected by the following pseudo expression: **[A-z.-+_]+([0-9][A-z]?)**

The **man2html** utility only recognizes lines with " **-** " (the normal separator between manpage references and summary text) while in keyword search mode.

The **man2html** utility can be hooked in a CGI script/program to convert manpages on the fly. This is the reason for the **-cgiurl** option.

LIMITATIONS

The order that section head mapping is searched is not defined. Therefore, if two or more **<section head text>** can match a give manpage section, there is no way to determine which map tag is chosen.

If **-seealso** is specified, all xrefs are detected after the **SEE ALSO** heading. In other words, sections after **SEE ALSO** may contain hyperlinked xrefs.

BUGS

Text that is flush to the left margin, but is not actually a section head, can be mistaken for a section head. This mistake is more likely when the **-sun** option is in affect.

VERSION

This documentation describes **man2html** version 3.1 (20220714)

SEE ALSO

man(1), **nroff**(1), **perl**(1)

https://invisible-island.net/scripts/man2html.html

AUTHOR

Earl Hood
Thomas E. Dickey

ERRORS AND OMISSIONS

Troff version of this document initially created for version 2.1.0 by C. Jeffery Small (*jeff@cjsa.com*) by copying, reformatting, rearranging and partially rewriting the contents of the ascii text file **doc/man2html.txt**.