NAME
>   man2html – convert a UNIX manual page file from nroff/troff -man format to HTML

SYNOPSIS
>   **man2html** [**−check-html**] [**−grammar-level** *grammar*] [**−outdirectory** *directoryname*]
>   [**−prettyprint**] [**−split-limit** *filesize-in-bytes*]
>   *input-manpage-file(s)*

DESCRIPTION
>   **man2html** converts UNIX manual page files named on the command line, from **nroff**(1)/**troff**(1) *−man*
>   format to strictly-grammar-conforming HTML.
>
>   The output files have the same base name (or the base name with a numeric suffix, if output HTML file
>   splitting is requested), but extension *.html*.
>
>   Although some vendors, such as Sun Microsystems, provide clear documentation of how manual pages
>   should be written, many manual page authors ignore those recommendations, and use arbitrary [**nt**]**roff**
>   markup to achieve the traditional appearance of UNIX manual pages, without actually using the standard
>   *−man* format commands.
>
>   **man2html** works quite well on Sun manual pages, but may be less successful on manual pages from other
>   sources. In such a case, an alternative may be to use T. A. Phelp's **RosettaMan**(1), commonly installed as
>   **rman**(1). That program works on the *output* of **nroff**(1), and attempts to guess manual page structure from
>   the horizontal and vertical spacing in order to add HTML markup. When vendor-provided manual pages
>   are available only in preformatted form, as on IBM AIX and SGI IRIX systems, **rman**(1) may be your only
>   choice. However, when **man2html** can be used successfully, it can often do a better job than **rman**(1),
>   because it has a better understanding of the document structure implied by [**nt**]**roff** manual-page markup.

OPTIONS
>   Command-line options may be abbreviated to any unique prefix, and letter case is *significant*. Options and
>   files are processed in the order found; thus, options affect only files that follow them on the command line.
>
>   **−check-html**   Check the output HTML for validity with a rigorous SGML parser, such as **html-check**(1)
>   or **html-ncheck**(1).
>
>   **−grammar-level** *grammar*
>   Specify a grammar level to select a suitable
>   `<!DOCTYPE HTML PUBLIC "...">`
>
>   declaration. Acceptable values are: *0*, *1*, *2*, *2-strict*, *3*, *3-strict*, *3.2*, *4*, *4-loose*, *Cougar*,
>   *Mosaic*, and *Netscape* [default: *2*].
>
>   **−outdirectory** *directoryname*
>   Provide an alternate directory into which output HTML files are placed. The default is the
>   current directory.
>
>   **−prettyprint**   Prettyprint each output file with **html-pretty**(1). Prettyprinting is done before any syntax
>   check requested by the **−check-html** option.
>
>   **−split-limit** *filesize-in-bytes*
>   Split the translated HTML from input files that are larger than the specified size into multi-
>   ple output HTML files: a root file, and section files named *basename-nn.html*, where *base-*
>   *name* is the manual page file name with directory path and extension removed, and *nn* is a
>   section number 01, 02, . . . [default: no output splitting].
>
>   The root file will contain a table of contents that directs the reader to the section files, and
>   each of those begins and ends with a navigation command area that allows moving one to
>   three sections in either direction, as well as back up to the root file.
>
>   This option permits large manual page files to be split into smaller parts that load faster
>   over the World-Wide Web, although with the possibly significant disadvantage that the
>   reader can no longer search the entire document with a single command.

HTML GRAMMAR LEVELS

The level 3 grammar has expired; some of its features, particularly the support for markup of mathematics, will appear in a future HTML grammar level.

The version 3.2 grammar is a stopgap, which, despite its higher number, lies approximately between 2 and 3 in features. It was released on November 5, 1996, at `http://www.w3.org/pub/WWW/` in order to provide a stable grammar toward which WWW browser developers could work.

The next version of HTML, code-named *Cougar*, is under development, and will become version 4.0 when it is finally released. The first draft public release was on 8 July 1997, and that was followed by a proposed recommended version on 7 November 1997.

There are only four potential differences in the output of **man2html** for these grammar levels:

- The output `<!DOCTYPE HTML PUBLIC "...">` declaration depends on the grammar level.

- At version 3 and above, the SGML entity * * can be used for non-breakable space instead of the less obvious numeric entity * * which is required by the level 2 grammar.

- At versions 3 and 3.2, the SGML entity *"*, representing a quotation mark, must be replaced by a numeric entity, *"*, because of an unfortunate error of omission in the grammars.

- At version 3.2 and higher, the output HTML will use *<CENTER> ... </CENTER>* directives to support centered text. At earlier grammar levels, centering requests are ignored, but the request is preserved in a comment, and lines are still broken as they would be when centered.

  Centering is exceedingly rare in manual page files (it is completely absent from *all* of Sun's standard manual pages), so the default level 2 grammar should almost always be sufficient.

SEE ALSO

amaya(1), **arena**(1), **chimera**(1), **grail**(1), **hotjava**(1), **html-check**(1), **html-ncheck**(1), **html-norm**(1), **html-pretty**(1), **html-spam**(1), **html2latex**(1), **htmlchek**(1), **jde**(1), **latex2html**(1), **lynx**(1), **netscape**(1), **nsgmls**(1), **panorama**(1), **rman**(1), **RosettaMan**(1), **rtf2html**(1), **sgmlnorm**(1), **sgmls**(1), **spam**(1), **spent**(1), **texi2html**(1), **xmosaic**(1).

AUTHOR

Nelson H. F. Beebe, Ph.D.
Center for Scientific Computing
University of Utah
Department of Mathematics, 322 INSCC
155 S 1400 E RM 233
Salt Lake City, UT 84112-0090
USA
Tel: +1 801 581 5254
FAX: +1 801 585 1640, +1 801 581 4148
Email: beebe@math.utah.edu, beebe@acm.org, beebe@ieee.org (Internet)
WWW URL: http://www.math.utah.edu/~beebe

AVAILABILITY

man2html is freely available; its master distribution can be found at

```
ftp://ftp.math.utah.edu/pub/sgml/
```

in the file *man2html-x.yy.tar.gz* where *x.yy* is the current version. Other distribution formats are usually available in the same location. Several other SGML and HTML tools are available in that same directory.

That site is mirrored to several other Internet archives, so you may also be able to find it elsewhere on the Internet; try searching for the string *man2html* at one or more of the popular Web search sites, such as

```
http://altavista.digital.com/
http://www.hotbot.com/
```

```
http://www.stpt.com/
http://www.yahoo.com/
```