

NAME

PolyglotMan, rman - reverse compile man pages from formatted form to a number of source formats

SYNOPSIS

rman [*options*] [*file*]

DESCRIPTION

Up-to-date instructions can be found at <http://polyglotman.sourceforge.net/rman.html>

PolyglotMan takes man pages from most of the popular flavors of UNIX and transforms them into any of a number of text source formats. PolyglotMan was formerly known as RosettaMan. The name of the binary is still called *rman*, for scripts that depend on that name; mnemonically, just think "reverse man". Previously *PolyglotMan* required pages to be formatted by nroff prior to its processing. With version 3.0, it *prefers* [tn]roff source and usually produces results that are better yet. And source processing is the only way to translate tables. Source format translation is not as mature as formatted, however, so try formatted translation as a backup.

In parsing [tn]roff source, one could implement an arbitrarily large subset of [tn]roff, which I did not and will not do, so the results can be off. I did implement a significant subset of those use in man pages, however, including tbl (but not eqn), if tests, and general macro definitions, so usually the results look great. If they don't, format the page with nroff before sending it to PolyglotMan. If PolyglotMan doesn't recognize a key macro used by a large class of pages, however, e-mail me the source and a uuencoded nroff-formatted page and I'll see what I can do. When running PolyglotMan with man page source that includes or redirects to other [tn]roff source using the .so (source or inclusion) macro, you should be in the parent directory of the page, since pages are written with this assumption. For example, if you are translating /usr/man/man1/lis.1, first cd into /usr/man.

PolyglotMan accepts man pages from: SunOS, Sun Solaris, Hewlett-Packard HP-UX, AT&T System V, OSF/1 aka Digital UNIX, DEC Ultrix, SGI IRIX, Linux, FreeBSD, SCO. Source processing works for: SunOS, Sun Solaris, Hewlett-Packard HP-UX, AT&T System V, OSF/1 aka Digital UNIX, DEC Ultrix. It can produce printable ASCII-only (control characters stripped), section headers-only, Tk, TkMan, [tn]roff (traditional man page source), SGML, HTML, MIME, LaTeX, LaTeX2e, RTF, Perl 5 POD. A modular architecture permits easy addition of additional output formats.

The latest version of PolyglotMan is available from <http://polyglotman.sourceforge.net/>.

OPTIONS

The following options should not be used with any others and exit PolyglotMan without processing any input.

-h|--help Show list of command line options and exit.

-v|--version Show version number and exit.

You should specify the filter first, as this sets a number of parameters, and then specify other options.

-f|--filter <ASCII|roff|TkMan|Tk|Sections|HTML|SGML|MIME|LaTeX|LaTeX2e|RTF|POD>
Set the output filter. Defaults to ASCII.

-S|--source PolyglotMan tries to automatically determine whether its input is source or formatted; use this option to declare source input.

-F|--format|--formatted
PolyglotMan tries to automatically determine whether its input is source or formatted; use this option to declare formatted input.

-l|--title *printf-string*
In HTML mode this sets the <TITLE> of the man pages, given the same parameters as *-r*.

-r|--reference|--manref *printf-string*
In HTML and SGML modes this sets the URL form by which to retrieve other man pages. The string can use two supplied parameters: the man page name and its section. (See the Examples section.) If the string is null (as if set from a shell by "-r """), '-' or 'off', then man page references will not be HREFs, just set in italics. If

your printf supports XPG3 positions specifier, this can be quite flexible.

-V|--volumes *<colon-separated list>*

Set the list of valid volumes to check against when looking for cross-references to other man pages. Defaults to *1:2:3:4:5:6:7:8:9:o:l:n:p* (volume names can be multi-character). If a non-whitespace string in the page is immediately followed by a left parenthesis, then one of the valid volumes, and ends with optional other characters and then a right parenthesis--then that string is reported as a reference to another manual page. If this -V string starts with an equals sign, then no optional characters are allowed between the match to the list of valids and the right parenthesis. (This option is needed for SCO UNIX.)

The following options apply only when formatted pages are given as input. They do not apply or are always handled correctly with the source.

- b|--subsections** Try to recognize subsection titles in addition to section titles. This can cause problems on some UNIX flavors.
- K|--nobreak** Indicate manual pages don't have page breaks, so don't look for footers and headers around them. (Older nroff -man macros always put in page breaks, but lately some vendors have realized that printout are made through troff, whereas nroff -man is used to format pages for reading on screen, and so have eliminated page breaks.) *PolyglotMan* usually gets this right even without this flag.
- k|--keep** Keep headers and footers, as a canonical report at the end of the page. *changeleft* Move changebars, such as those found in the Tcl/Tk manual pages, to the left. --> *notaggressive Disable* aggressive man page parsing. Aggressive manual, which is on by default, page parsing elides headers and footers, identifies sections and more. -->
- n|--name** *name* Set name of man page (used in roff format). If the filename is given in the form "*name . section*", the name and section are automatically determined. If the page is being parsed from [tn]roff source and it has a .TH line, this information is extracted from that line.
- p|--paragraph** paragraph mode toggle. The filter determines whether lines should be linebroken as they were by nroff, or whether lines should be flowed together into paragraphs. Mainly for internal use.
- s|section #** Set volume (aka section) number of man page (used in roff format). *tables* Turn on aggressive table parsing. -->
- t|--tabstops #** For those macros sets that use tabs in place of spaces where possible in order to reduce the number of characters used, set tabstops every # columns. Defaults to 8.

NOTES ON FILTER TYPES

ROFF

Some flavors of UNIX ship man page without [tn]roff source, making one's laser printer little more than a laser-powered daisy wheel. This filter tries to intuit the original [tn]roff directives, which can then be recompiled by [tn]roff.

TkMan

TkMan, a hypertext man page browser, uses *PolyglotMan* to show man pages without the (usually) useless headers and footers on each pages. It also collects section and (optionally) subsection heads for direct access from a pulldown menu. TkMan and Tcl/Tk, the toolkit in which it's written, are available via anonymous ftp from <ftp://ftp.smli.com/pub/tcl/>

Tk

This option outputs the text in a series of Tcl lists consisting of text-tags pairs, where tag names roughly correspond to HTML. This output can be inserted into a Tk text widget by doing an *eval <text-widget> insert end <text>*. This format should be relatively easily parsable by other programs that want both the text and the tags. Also see ASCII.

ASCII

When printed on a line printer, man pages try to produce special text effects by overstriking characters with themselves (to produce bold) and underscores (underlining). Other text processing software, such

as text editors, searchers, and indexers, must counteract this. The ASCII filter strips away this formatting. Piping `nroff` output through `col -b` also strips away this formatting, but it leaves behind unsightly page headers and footers. Also see Tk.

Sections

Dumps section and (optionally) subsection titles. This might be useful for another program that processes man pages.

HTML

With a simple extension to an HTTP server for Mosaic or other World Wide Web browser, *PolyglotMan* can produce high quality HTML on the fly. Several such extensions and pointers to several others are included in *PolyglotMan*'s *contrib* directory.

SGML

This is approaching the Docbook DTD, but I'm hoping that someone that someone with a real interest in this will polish the tags generated. Try it to see how close the tags are now.

MIME

MIME (Multipurpose Internet Mail Extensions) as defined by RFC 1563, good for consumption by MIME-aware e-mailers or as Emacs (>=19.29) enriched documents.

LaTeX and LaTeX2e

Why not?

RTF

Use output on Mac or NeXT or whatever. Maybe take random man pages and integrate with NeXT's documentation system better. Maybe NeXT has own man page macros that do this.

PostScript and FrameMaker

To produce PostScript, use *groff* or *psroff*. To produce FrameMaker MIF, use FrameMaker's builtin filter. In both cases you need *[tn]roff* source, so if you only have a formatted version of the manual page, use *PolyglotMan*'s *roff* filter first.

EXAMPLES

To convert the *formatted* man page named *ls.1* back into *[tn]roff* source form:

```
rman -f roff /usr/local/man/cat1/ls.1 > /usr/local/man/man1/ls.1
```

Long man pages are often compressed to conserve space (compression is especially effective on formatted man pages as many of the characters are spaces). As it is a long man page, it probably has subsections, which we try to separate out (some macro sets don't distinguish subsections well enough for *PolyglotMan* to detect them). Let's convert this to LaTeX format:

```
pcat /usr/catman/a_man/cat1/automount.z | rman -b -n automount -s 1 -f latex > automount.man
```

Alternatively, *man 1 automount* | *rman -b -n automount -s 1 -f latex > automount.man*

For HTML/Mosaic users, *PolyglotMan* can, without modification of the source code, produce HTML links that point to other HTML man pages either pregenerated or generated on the fly. First let's assume pregenerated HTML versions of man pages stored in */usr/man/html*. Generate these one-by-one with the following form:

```
rman -f html -r 'http:/usr/man/html/%s.%s.html' /usr/man/cat1/ls.1 > /usr/man/html/ls.1.html
```

If you've extended your HTML client to generate HTML on the fly you should use something like:

```
rman -f html -r 'http://bin/man2html?%s:%s' /usr/man/cat1/ls.1
```

when generating HTML.

BUGS/INCOMPATIBILITIES

PolyglotMan is not perfect in all cases, but it usually does a good job, and in any case reduces the problem of converting man pages to light editing.

Tables in formatted pages, especially H-P's, aren't handled very well. Be sure to pass in source for the page to recognize tables.

The man pager *woman* applies its own idea of formatting for man pages, which can confuse *PolyglotMan*. Bypass *woman* by passing the formatted manual page text directly into *PolyglotMan*.

The *[tn]roff* output format uses *fB* to turn on boldface. If your macro set requires *.B*, you'll have to a postprocess the *PolyglotMan* output.

SEE ALSO

tkman(1) , *xman(1)* , *man(1)* , *man(7)* or *man(5)* depending on your flavor of UNIX

AUTHOR

PolyglotMan

by Thomas A. Phelps (*phelps@ACM.org*)

developed at the

University of California, Berkeley

Computer Science Division

Manual page last updated on \$Date: 1998/07/13 09:47:28 \$