
Man Pages as HTML is hard

And one Dollar per request could be saved.

Onkobu Tanaake

2018-09-20T10:00:00

Various tools exist and have existed even before I had internet at all. One of the earliest still in use is man2html. A more recent variant is roffit. I tried both and even attempted to improve them – I gave up.

My average desktop Linux offers nearly 3000 level 1 man pages. They are full of sections and program arguments. But man format (man 7 man) is not a very good start for structured documents. Roffit's sample of curl man page is very good but also made for each other. Running roffit for xinput's man page yields no anchors for the program arguments. Translating zstdcat's man page with roffit even causes partly garbled section headings and arguments.

Some of the issues available:

- Empty lines expressed as lonely period/ ''
- Inconsistent use of paragraphs .P, .PP, .LP
- Inconsistent rendering of indentation, e.g. zstdcat section *DICTIONARY BUILDER* with roffit
- Formatting only omitting any context

The last point is the most serious. There is no outline of program arguments. Instead a man page author could use a new paragraph with a bold marker for the argument itself and text afterwards:

```
.PP
.TP 8
.B --enable \fIdevice\fP
Enable the \fIdevice\fP. This call is equivalent to
.B xinput --set-prop device \fI"Device Enabled"\fP 1
```

A more experienced author would use a indexed paragraph that allows a sort of heading or title. The next sample is taken from curl's man page:

```
.IP "-v, --verbose"
Makes curl verbose during the operation. Useful for debugging and
what's going on "under the hood". A line starting with '>' means "
sent by curl, '<' means "header data" received by curl that is hid
normal cases, and a line starting with '*' means additional info p
curl.
```

Holding the previous two against the man page of `zstdcat` illustrates a lack of standardization for arguments at all. This third man page uses not only the character '#' but also escapes minus signs with backslashes:

```
.  
.TP  
\fB\-\-stream\-size=#\fR  
Sets the pledged source size of input coming from a stream\.
```

Rendering man pages with anchors for program arguments requires sophisticated pattern matching. It is not only hard to find arguments but also to transform them into a valid HTML-anchor.

And all this is already summarized in the man page level 7 of `man` itself:

...even HTML has more semantic markings). This situation makes it harder to vary the man format for different media, to make the formatting consistent for a given media, and to automatically insert cross-references.

So instead use the already available info pages created from TeXinfo. Write a sane primer as man page and the full version only as info page. In turn there is less post production effort when rendering for other media. There is not only `@option` but also `@opindex`, `@defun` and many more to add meaning to arguments.