# Exiv2 - Feature #1165

## Exiv2 man page structural enhancements

02 Mar 2016 20:16 - Robin Mills

| | | | | |
|---|---|---|---|---|
| **Status:** | Assigned | | **Start date:** | 02 Mar 2016 |
| **Priority:** | Normal | | **Due date:** | |
| **Assignee:** | Robin Mills | | **% Done:** | 20% |
| **Category:** | documentation | | **Estimated time:** | 4.00 hours |
| **Target version:** | 0.28 | | | |

**Description**

I have received the following email from Eric Raymond:

> The exiv2 man page needs some serious restructuring. The problems
> aren't obvious if you think of it as purely visual markup, but they
> interfere with structural translation to other formats like DocBook
> XML.
>
> I maintain a utility called doclifter that does that. It handles about
> 96% of manual pages automatically. Most of the rest need only a point
> patch or two to be ready. Due to frequent use of very low-level
> constructs like .br, the exiv2 page is one of the exceptions. It
> needs a pretty extensive rework.
>
> I need to know you are interested before writing patches that large.
> I may also have some questions about the intention of the documentation.

**Related issues:**

| | | |
|---|---|---|
| Related to Exiv2 - Feature #1061: Exiv2 manpage overhaul | **Assigned** | **22 Apr 2015** |

## Associated revisions

**Revision 4209 - 02 Mar 2016 20:18 - Robin Mills**

#1165. Thanks to Eric for reporting this and providing the patch.

## History

**#1 - 02 Mar 2016 20:17 - Robin Mills**

*- File exiv2.1.patch added*

---

After further discussion:
doclifter can be used as a lint/validator, but that's not its main purpose.
Its main purpose is to lift troff documents to XML-DocBook, a structure-centric
document format which, among other things, can be rendered to better HTML than
you could get from a naive troff-to-HTML translation.

To do its work, doclifter recognizes that although man page markup is
visual, it is almost always used in stereotyped ways that imply
document structure. It applies hundreds of rules like the following:
"If the current section name is FILES, and you see a bolded token
containing slashes, that is a filename and the visual bolding
should become a <filename> tag in XML."

Thus, you feed in unstructured troff tag soup and get out a structured
XML document full of semantic hints. Better for making HTML, better
for searching. My long term goal is to get the entire Linux/*BSD
manpage corpus to the point where it is XML-conversion-ready so that
distributions can (a) start presenting documentation as an cross-linked
set of HTML pages, (b) build semantic-search tools that mine the XML.

I've gotten surprisingly far with this in just twelve years. 96% of the

manual pages in a typical Linux distro lift without fuss (when I started
it was around 75%). Every year or two I go through the entire set in whatever
distro I'm using, make a set of patches to clean up as much as I can, then
ship them. In another two to four years I think I will be effectively done,
except for people randomly introducing errors when they modify things. I
have, for example, cleaned up the **entire** set of X pages.

The whole plan depends on most pages mostly using high-level markup -
.SH and .PP and friends rather than .br and .in and so forth, TBL
rather than tabular displays made with .nf/fi, that sort of thing.

The exiv2 page uses an unusually large amount of low-level markup.
For example, this:

```
.TP
.B \-p \fImode\fP
Print mode for the 'print' action. Possible modes are:
.br
s : print a summary of the Exif metadata (the default)
.br
a : print Exif, IPTC and XMP metadata (shortcut for \-Pkyct)
.br
t : interpreted (translated) Exif tags (\-PEkyct)
.br
v : plain Exif tag values (\-PExgnycv)
.br
h : hexdump of the Exif data (\-PExgnycsh)
.br
i : IPTC datasets (\-PIkyct)
.br
x : XMP properties (\-PXkyct)
.br
c : JPEG comment
.br
p : list available image previews, sorted by preview image size in pixels
.br
S : print image structure information (jpg, png and tiff only)
.br
X : print "raw" XMP (jpg, png and tiff only)
```

From a structural-translation point of view this is awful - you just can't
do anything with those .br tags. This ought to be expressed as a table
or a .TP list, which docliter can recognize as structure and turn into an
XML table - from which you get an HTML table when you render to HTML.

You are correct that no actual content changes would be required. I just
want to fix the markup.

--------------------------
An a patch (attached).

Some explanation:

I turned a couple of mode displays made with lines separated by .br into tables.
A good general rule is to never use .br at all. If you want to, say, continue
a multi-paragraph .TP body, .sp 1 will do that.

I wrapped code and command examples in this:

```
.RS
.nf
<example goes here>
.fi
.RE
```

That is a cliche that doclifter can recognize and run into a <literallayout>
display. The .RS/.RE pair is a less fragile way of saying "Indent this" than
a literal indent.

There were a few places where you had constructions like this:

```
.TP
.nf
exiv2 \-M"set Exif.Photo.UserComment charset=Ascii New Exif comment" image.jpg
.fi
```

```
Sets the Exif comment to an ASCII string.
```

I removed the .nf/.fi pair in those cases. I realize this may cause line filling where it's not wanted, but the way it was written is pretty much guaranteed to choke any semantic parser and any viewer other than groff itself.

The problem is, most viewers have hardwired in the assumption that the very first line after the .TP is the hanging label. The .nf there will cause mass confusion and the rest of the display will be misinterpreted.

The right thing to do would be to rwrite these examples so they aren't as long, but I don't have the domain knowledge to do that.

Thanks for your cooperation.  If you have any other questions I will be happy to answer them.


**#2 - 02 Mar 2016 20:20 - Robin Mills**

*- % Done changed from 50 to 20*

*- Estimated time changed from 2.00 h to 4.00 h*


Patch submitted [r4209](r4209).  Thank You to Eric for reporting this and providing the patch.

I will update our test suite to extract and test the code snippets in the man page.


**#3 - 13 Jun 2016 06:21 - Robin Mills**

*- Target version changed from 0.26 to 0.28*


Deferred for v0.27 to make the schedule.


**Files**

| | | | | |
|---|---|---|---|---|
| exiv2.1.patch | 6.93 KB | 02 Mar 2016 | | Robin Mills |