How can I diff two XML files?

On Linux, how could I generate a diff between two XML files?

Ideally, I would like to be able configure it to some things strict, or loosen some things, like whitespace, or attribute order.

I'll often care that the files are functionally the same, but diff by itself, would be annoying to use, especially if the XML file doesn't have a lot of linebreaks.

For example, the following should really be okay to me:

```
<tag att1="one" att2="two">
  content
</tag>

<tag att2="two" att1="one">
  content
</tag>
```

linux xml diff

edited Aug 8 '12 at 10:27
slhck
125k ● 38 ● 310 ● 359

asked Dec 7 '09 at 16:27
qedi
651 ● 1 ● 6 ● 10

6 Answers

One approach would be to first turn both XML files into Canonical XML, and compare the results using `diff`. For example, xmllint can be used to canonicalize XML.

```
$ xmllint --c14n one.xml > 1.xml
$ xmllint --c14n two.xml > 2.xml
$ diff 1.xml 2.xml
```

answered Dec 9 '09 at 20:06
Jukka Matilainen
1,494 ● 10 ● 6

1 Never knew about the --c14n switch in xmllint. That's handy. – qedi Dec 10 '09 at 20:21

9 You can do it in one line too `vimdiff <(xmllint --c14n one.xml) <(xmllint --c14n two.xml)` – Nathan Villaescusa Mar 3 '13 at 1:53

and xmllint ships with OS X – ClintM Sep 20 at 16:17

Tried to use @Jukka Matilainen's answer but had problems with white-space (one of the files was a huge one-liner). Using `--format` helps to skip white-space differences.

```
xmllint --format one.xml > 1.xml
xmllint --format two.xml > 2.xml
diff 1.xml 2.xml
```

Note: Use `vimdiff` command for side-by-side comparison of the xmls.

edited Aug 8 '12 at 10:33

answered Aug 8 '12 at 9:58
GuruM
191 ● 1 ● 5

In my case `two.xml` was generated from `one.xml` by a script. So I just needed to check what was added/removed by the script. – GuruM Aug 8 '12 at 10:36

This was the option I needed. Supposedly the most canonical version can be obtained by combining `--format` with `--exc-c14n`; will probably be still slower to process :(– VINCENT Nov 27 '14 at 14:05

It's been quite some time since I wrote the answer, but I faintly remember using the --exc-c14n flag. However, diff-ing the output with/without the flag showed no differences so just stopped using it. Dropping unnecessary/unused flags might make the process faster. – GuruM Dec 21 '14 at 6:49

2 The `--exc-c14n` option specifies sorting of the attributes. In your specific files the attributes probably were already sorted, but the general advice would be to use the combination `--format --exc-c14n`. – VINCENT Dec 22 '14 at 14:33

Jukka's answer did not work for me, but it did point to Canonical XML. Neither *--c14n* nor *--c14n11* sorted the attributes, but i did find the ***--exc-c14n*** switch did sort the attributes. ***--exc-c14n*** is not listed in the man page, but described on the command line as "W3C exclusive canonical format".

```
$ xmllint --exc-c14n one.xml > 1.xml
$ xmllint --exc-c14n two.xml > 2.xml
$ diff 1.xml 2.xml

$ xmllint | grep c14
    --c14n : save in W3C canonical format v1.0 (with comments)
    --c14n11 : save in W3C canonical format v1.1 (with comments)
    --exc-c14n : save in W3C exclusive canonical format (with comments)

$ rpm -qf /usr/bin/xmllint
libxml2-2.7.6-14.el6.x86_64
libxml2-2.7.6-14.el6.i686

$ cat /etc/system-release
CentOS release 6.5 (Final)
```

Warning ***--exc-c14n*** strips out the xml header whereas the --c14n prepends the xml header if not there.

answered Mar 4 '14 at 12:51
rjt
386 ● 1 ● 5 ● 11

Diffxml gets the basic functionality correct, though it doesn't seem to offer many options for configuration.

answered Dec 7 '09 at 16:57
dsolimano
2,322 ● 2 ● 12 ● 30

It's not quite there yet, but it looks promising at least. – qedi Dec 7 '09 at 17:02

I use Beyond Compare to compare all types of text based files. They produce versions for Windows and Linux.

answered Dec 7 '09 at 16:30
Alan
153 ● 1 ● 7

Plain text comparisons would say the two lines differed, whereas the OP wants them to be reported as the same. – ChrisF Dec 7 '09 at 16:33

2 i.e. **Canonically compare** the XML. – Chris W. Rea Dec 9 '09 at 20:08

Beyond Compare really sucks for this. It seems to just not be aware of XML elements and do mostly just text comparison. – Rob K May 23 at 17:54

Our SD Smart Differencer compares documents based on *structure* as opposed to actual layout.

There's an XML Smart Differencer. For XML, that means matching order of tags and content. It should note that the text string in the specific fragment you indicated was different. It presently doesn't understand the XML notion of tag attributes indicating whether whitespace is normalized vs. significant.

edited Nov 27 '14 at 15:52

answered May 23 '10 at 5:01
Ira Baxter
106 ● 1 ● 2 ● 16

In your SO profile you provide full disclosure about your employer; I'd have preferred a short disclaimer inside your answer as well :) BTW, I tried to download an evaluation copy, but the request form is 'smart' (via JS) enough to disable the combination XML with Smart Differencer (also the latter in combination with Python, although possible according to the SD product page)? – VINCENT Nov 27 '14 at 14:03

1 Ah. Thanks for the reminder. This is an answer from a time before there was a clear SO policy on this. I'm revising the answer to signal the relationship in SO policy compliant answer. – Ira Baxter Nov 27 '14 at 15:52

I'll check the download page; not all of our live products make into that list. Yes, these exist. – Ira Baxter Nov 27 '14 at 15:53

I checked the download page. Yes, the XML smart differencer is not there. I'll have the back-room guys work on fixing that; should be there in 1-2 weeks at most (they have a backlog, don't we all?) In meantime, if you want to try it, send email (see bio). – Ira Baxter Jan 26 '15 at 20:16

Thanks, but I'm patient :) – VINCENT Jan 26 '15 at 22:31