

RFC 5261 : An Extensible Markup Language (XML) Patch Operations Framework Utilizing XML Path Language (XPath) Selectors

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 11 septembre 2008

Date de publication du RFC : Septembre 2008

<https://www.bortzmeyer.org/5261.html>

De nombreux protocoles IETF s'appuient sur XML, par exemple pour stocker de l'information de configuration. Dans ce cadre, le besoin de modifier de l'information XML sans devoir transporter le fichier entier, en communiquant juste les différences, est fréquent.

Pour le texte brut, ce besoin est couvert depuis longtemps par le format diff / patch. Jusqu'à présent, il n'existait aucune norme équivalente pour XML. Plusieurs propositions de « langages de "patch" » pour XML ont été faites <<https://www.bortzmeyer.org/langages-patch-xml.html>> mais ce RFC est le premier où on atteint le stade de la normalisation.

Il a été développé au sein du groupe de travail Simple <<http://www.ietf.org/html.charters/simple-charter.html>>, qui normalise des extensions au protocole SIP (RFC 3261¹) et qui en avait besoin, entre autres, pour la mise à jour de configurations stockées en XCAP (RFC 4825). Mais il est très général et applicable à tous les domaines.

Le principe de cette technique est simple. La partie du document XML à modifier est identifiée par une expression XPath. Une fois cette partie localisée, du contenu est ajouté, modifié ou retiré, selon que le fichier de modifications contienne des éléments <add>, <replace> ou <remove> (la section 4 détaille chacun de ces éléments).

L'annexe A du RFC contient de nombreux exemples, en voici quelques uns pour illustrer :

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc3261.txt>

```
<diff>
  <add sel="/section"><para>Nouveau paragraphe, lorem, ipsum, dolor, etc</add>
</diff>
```

Ce code ci-dessus ajoute un nouvel élément `<para>`. L'expression XPath pour localiser le point où l'ajouter est donnée par l'attribut `sel`. Ici, on ajoute le nouvel élément dans l'élément `<section>` de premier niveau.

```
<diff>
  <replace sel="section[title='Bibliographie']">
    <bibliography>...</bibliography>
  </replace>
</diff>
```

Ici, la section de titre « Bibliographie » est remplacée par un élément `<bibliography>`.

```
<diff>
  <remove sel="article/section/para/text()"/>
</diff>
```

Ici, le texte de l'élément `<para>` est supprimé.

La syntaxe complète de ce nouveau format est décrite dans la section 8, elle est exprimée en W3C Schema.

À noter qu'il existe déjà une mise en œuvre de ce format, `xmlpatch` <http://sourceforge.net/projects/xmlpatch>. Voici un exemple d'utilisation avec un premier document `toto.xml` :

```
<article>
<abstract>Brouillon</abstract>
<section>
<title>Premier essai</title>
<para>À l'heure actuelle, on ne sait pas encore grand'chose.</para>
</section>
</article>
```

et sa deuxième version, `totoplus.xml` :

```
<article>
<section>
<title>Premier essai</title>
<para>Aujourd'hui, nous en savons bien plus.</para>
</section>
<section>
<title>Révélations</title>
<para>Nous avons même une seconde section.</para>
</section>
</article>
```

Appliquons `xml_diff` pour obtenir un fichier de différences, au format du RFC :

```
% xml_diff -f toto.xml -t totoplus.xml -o toto.xmlpatch
```

Ce fichier des différences contient :

```
<?xml version="1.0"?>
<x:changes xmlns:x="urn:xml-changes">
  <x:remove sel="*/*[1]" ws="after"/>
  <x:replace sel="*/*[2]/text()">Aujourd'hui, nous en savons bien plus.</x:replace>
  <x:add sel="*"><section>
<title>R&#xE9;v&#xE9;lutions</title>
<para>Nous avons m&#xEA;me une seconde section.</para>
</section>
</x:add></x:changes>
```

Personnellement, je ne trouve pas les expressions XPath extraordinaires, elles manquent notamment de robustesse en cas de légers changements du document source. Mais elles sont correctes et on peut appliquer ce *"patch"* :

```
% xml_patch -f toto.xml -p toto.xmlpatch -o toto-patched.xml
```

et cela nous redonne un document qui a le même contenu que `totoplus.xml`. Attention, cela ne signifie pas qu'il soit identique au bit près (par exemple, les caractères UTF-8 de l'original ont été remplacés par des **entités** numériques XML) mais que son contenu informationnel (l'*"infoset"* XML) est le même.

Au passage, vous remarquerez que l'élément XML qui englobe les éléments *"patch"* n'est pas défini par le RFC. `<diff>` (dans les exemples du RFC) ou bien `<changes>` (dans l'exemple avec `xml_patch`) sont des choix arbitraires. Cela été fait ultérieurement dans le RFC 7351.