

RFC 7998 : "xml2rfc" Version 3 Preparation Tool Description

Stéphane Bortzmeyer
<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 16 décembre 2016

Date de publication du RFC : Décembre 2016

<https://www.bortzmeyer.org/7998.html>

Depuis la sortie du RFC 7990¹, le format canonique des RFC est le format XML. C'est un texte écrit en XML, selon le vocabulaire du RFC 7991 qui sera la version de référence, archivée et faisant autorité, pour tous les RFC. Comme les auteurs enverront souvent un XML imparfait, un **outil de préparation** sera nécessaire, pour compléter ce XML imparfait et permettre au "*RFC Editor*" de travailler sur une base sérieuse. Ce nouveau RFC décrit ce futur outil de préparation.

Ainsi, les outils qui travailleront sur le format canonique des RFC (par exemple les outils qui produiront le PDF, le HTML, etc) pourront compter sur un document complet et n'auront pas à gérer les écarts des auteurs : seul l'outil de préparation devra s'en soucier.

Cet outil de préparation servira aux RFC une fois qu'ils seront soumis au "*RFC production center*" (cf. RFC 8728) mais aussi aux "*Internet-Drafts*" pendant leur élaboration.

Dans le second cas (section 3 de notre RFC), le futur outil de préparation prendra un "*Internet-Draft*" en entrée et produira un document complet (par exemple avec addition des "*boilerplates*").

Et ce sera à peu près la même chose lorsque le RFC sera presque fini. On passera la version finale dans l'outil de préparation, qui résoudra les références externes et ajoutera les éléments obligatoires manquants.

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc7990.txt>

Bon, et en quoi vont consister exactement ces modifications? Elle sont décrites dans la section 5, qui forme le gros de ce RFC. Contrairement à l'outil actuel `idnits` <<https://tools.ietf.org/tools/idnits/>> qui se contente de vérifier les "*Internet-Drafts*", le nouvel outil va également corriger le texte, ajoutant des éléments, et changeant les valeurs erronées.

C'est ainsi que l'outil de préparation va traiter les éléments `XInclude`, les remplaçant par la valeur incluse. Il va traiter les DTD pour les supprimer ensuite (donc, remplacer les entités par leur valeur, et inclure les fichiers inclus par ce mécanisme). Ces deux actions peuvent aujourd'hui être faites par l'outil `xmllint`, avec `xmllint --xinclude --noent --dropdtd NOMDUFICHIER.xml`.

Outre ces mécanismes d'inclusion de XML, l'outil de préparation va aussi traiter les inclusions spécifiques au vocabulaire du RFC 7991. Ainsi, `<artwork>` a un attribut `src` indiquant la source du graphique, et l'outil de préparation va donc inclure ce graphique. (Idem avec `<sourcecode>` pour inclure le code source.)

Les instructions XML (PI, "*Processing Instructions*") seront supprimées (ça, je ne sais pas le faire avec `xmllint`).

L'outil va valider le résultat produit, en utilisant la grammaire Relax NG du RFC 7991. Ça peut aujourd'hui se faire avec `xmllint` mais aussi avec `rnv` <<http://www.davidashen.net/rnv.html>> :

```
% rnv rfc-v3.rnc rfc-v3-sample.xml
rfc-v3-sample.xml
```

ou bien avec `jing` <<http://www.thaiopensource.com/relaxng/jing.html>> :

```
% java -jar ./jing-20091111/bin/jing.jar -c rfc-v3.rnc rfc-v3-sample.xml
```

Parmi les nombreuses transformations possibles, citons l'ajout (s'il n'était pas déjà présent) de l'élément `<seriesInfo>` qui indique s'il s'agit d'un "*Internet-Draft*" ou d'un RFC, l'ajout d'un élément `<date>` s'il manque (sa valeur étant la date du traitement), changement de l'ancien attribut `title` en `name`, le retrait des commentaires XML...

Il est fréquent dans les "*Internet-Drafts*" de voir des paragraphes qui ne devront pas être inclus dans le futur RFC. C'est le cas s'ils contiennent des exemples réels qui risquent de ne pas être éternels (les RFC peuvent durer longtemps et ne sont jamais modifiés). C'est également le cas s'il s'agit de l'état actuel des mises en œuvre d'un RFC, comme décrit dans le RFC 7942. Dans le système actuel, ces paragraphes sont marqués par un texte en langue naturelle. Dans le nouveau vocabulaire du RFC 7991, ce sera fait avec un attribut `removeInRFC`. L'outil de préparation pourra enlever automatiquement ce paragraphe quand il préparera un RFC.

L'outil de préparation devra également arranger le XML. Cela peut se faire aujourd'hui avec `xmllint` (ses options `--format` ou bien `--pretty`). Par contre, il n'est pas prévu de mettre le XML sous sa forme canonique.

Il y aura d'autres opérations faites par l'outil de préparation, voir le RFC pour les détails.

L'outil n'est pas encore développé, un appel d'offres a été lancé et les gagnants ont été les sociétés SeanTek et Elf Tools.