

RFC 5005 : Feed Paging and Archiving

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 8 septembre 2007

Date de publication du RFC : Septembre 2007

<https://www.bortzmeyer.org/5005.html>

Dans la désormais longue série des RFC autour du format Atom, celui-ci spécifie un moyen de récupérer une partie d'un flux de syndication et comment itérer pour avoir le reste du flux.

Atom, normalisé dans le RFC 4287¹, permet de décrire un flux de syndication, comprenant un certain nombre d'entrées, par exemple des articles d'un blog. Avant ce nouveau RFC, Atom ne permettait pas de récupérer le flux en plusieurs étapes. La plupart des sites Web ne publiaient que les N derniers articles et sans fournir de moyen d'obtenir les autres. Notre RFC change cela : il permet de récupérer des flux par étapes successives ("*pages feeds*") et aussi de récupérer un flux en plusieurs étapes en ayant la garantie que cette récupération inclut toutes les entrées ("*archived feeds*").

Un "*paged feed*" est simplement un flux qui contient des liens vers le groupe suivant d'entrées. Par exemple :

```
<feed xmlns="http://www.w3.org/2005/Atom">
<title>Mon flux à moi</title>
<link href="http://example.org/" />
<link rel="self" href="http://example.org/index.atom" />
<link rel="next" href="http://example.org/index.atom?page=2" />
...
```

1. Pour voir le RFC de numéro NNN, <https://www.ietf.org/rfc/rfcNNN.txt>, par exemple <https://www.ietf.org/rfc/rfc4287.txt>

où ce document, `index.atom`, contient un lien de type `next` qui mène au document suivant du flux, `index.atom?page=2`.

Ce type `next`, ainsi que d'autres types décrits dans ce RFC comme `prev` sont enregistrés dans le registre IANA des relations `<https://www.iana.org/assignments/link-relations.html>`.

Un "*paged feed*" n'offre pas de garantie de cohérence. Entre la requête à `index.atom` et une requête ultérieure à `index.atom?page=2`, des entrées ont pu être ajoutées et des entrées figurant dans le premier document ont pu « migrer » vers le second. Itérer sur les `next` peut donc ne donner qu'une partie de la collection d'entrées.

Pour avoir un flux cohérent, il faut utiliser les "*archived feeds*" avec les relations `next-archive` et `prev-archive` :

```
<feed xmlns="http://www.w3.org/2005/Atom"
xmlns:fh="http://purl.org/syndication/history/1.0">
<!-- L'espace de noms http://purl.org/syndication/history/1.0 est
défini dans ce RFC -->
<title>Mon flux à moi</title>
<link rel="current" href="http://example.org/index.atom"/>
<link rel="self" href="http://example.org/2003/11/index.atom"/>
<fh:archive/>
<link rel="prev-archive"
href="http://example.org/2003/10/index.atom"/>
...
```

Itérer sur `prev-archive` et `next-archive`, si ces relations sont présentes, est sûr et garanti de ne perdre aucune entrée.

Notre RFC définit également un troisième type de flux, les "*complete feeds*" où la présence de l'élément `<fh:complete/>` garantit que le flux contient toutes les entrées de la collection.

Désolé, mais ce blog ne met pas encore en œuvre ce RFC et ne fournit donc qu'un flux limité, réduit aux N derniers articles.