

# Sélection d'articles dans un flux de syndication

Stéphane Bortzmeyer

<stephane+blog@bortzmeyer.org>

Première rédaction de cet article le 3 avril 2008. Dernière mise à jour le 6 avril 2008

<https://www.bortzmeyer.org/filtres-syndication.html>

---

Ce serait bien d'avoir des sélections de mon flux de syndication (<<http://www.bortzmeyer.org/feed.atom>>) par sujet (PostgreSQL, XSLT, Python, etc). Existe t-il des services en ligne qui assurent ce genre de sélections ?

Comme je maîtrise complètement le processus de production du flux <<https://www.bortzmeyer.org/blog-implementation.html>>, je pourrais bien sûr générer N flux selon le sujet mais cela représente du travail, alors que les critères de sélection évoluent tout le temps.

Je me demande s'il n'existe pas plutôt un service tiers de sélection à qui on donnerait l'URL du flux, un critère de sélection (mot-clé, expression rationnelle, expression XPath) et qui renverrait un URL que je pourrais donner à des sites d'agrégation comme <<http://planete.postgresql.fr/>> ou comme <<http://fr.incubator.wikiblogplanet.com/>>, qui n'acceptent que les articles sur un certain sujet (PostgreSQL dans le premier cas, Wikipédia dans le second).

J'ai testé trois de ces services mais d'autres adresses seraient bienvenues.

Le premier est <<http://www.filtermyrss.com/>>. Il marche uniquement avec RSS, alors que mon flux est en Atom (RFC 4287<sup>1</sup>). Il ne filtre que par mot-clés, pouvant apparemment porter sur certains éléments (titre uniquement, par exemple).

Puis j'ai testé <<http://feedrinse.com/>>. Il nécessite une inscription, certes gratuite mais c'est quand même pénible. Après inscription, je dois être bête, je n'ai jamais pu le faire marcher.

Le service de référence le plus souvent cité pour ce genre d'opérations est Yahoo Pipes <<http://pipes.yahoo.com/>>, un mécanisme de programmation graphique, permettant de définir des traitements sur des flux d'information. Yahoo Pipes nécessite apparemment un compte Yahoo et Yahoo me

---

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

refuse l'inscription ("*This is not a valid email address*", dit-il, et mon identifiant OpenID ne suffit apparemment pas donc j'ai abandonné tout espoir d'utiliser Yahoo). Un problème analogue m'empêche de tester Fwicky <<http://www.fwicky.com/>> qui en outre se vante stupidement de ses brevets dès la page d'accueil.

Enfin, <<http://re.rephrase.net/filter/>> est celui que j'ai choisi. Il n'y a pas besoin d'inscription, il marche bien avec Atom. Les recherches se font par mot-clés, portant sur le texte ou sur une partie du texte (une documentation très complète est disponible <<http://re.rephrase.net/filter/?p=filters>>). En outre, le source du service est distribué par l'auteur.

C'est donc celui que j'utilise en ce moment. Par exemple <<http://re.rephrase.net/filter/?feed=http%3A%2F%2Fwww.bortzmeyer.org%2Ffeed.atom&filter=title%3Apostgresql+OR+title%3ASQL>> sélectionne tous les articles du flux qui ont « postgresql » ou « SQL » dans leur titre.

Les caractères composés du français nécessitent un peu de bricolage. Rephrase, écrit en PHP, n'a pas de support d'Unicode et fait une recherche « bête » en comparant les octets. Si le flux de syndication est en UTF-8, il faut mettre de l'UTF-8 dans le critère de tri. Ainsi, pour chercher tous les articles parlant de Wikipédia, l'URL sera <http://re.rephrase.net/filter/?feed=http%3A%2F%2Fwww.bortzmeyer.org%2Ffeed.atom&filter=title%3AWikip%C3%A9dia> puisque le caractère « é » se code par 0xC3 et 0xA9 en UTF-8.

Depuis ce choix, mon blog a été muni d'un moteur de recherche <<https://www.bortzmeyer.org/moteur-recherche-wsgi.html>>. Donc, une autre solution, à explorer, serait d'utiliser le moteur de recherche </search> pour produire des flux Atom dynamiques, avec des requêtes comme . Toutefois, le moteur de recherche de ce blog ne renvoyant que les N articles les plus pertinents, ce n'est pas actuellement idéal pour faire une vue d'un flux de syndication.

Merci à Benoit Deuffic, David Larlet et Olivier Mengué pour leur aide à trouver ces services. Merci à l'auteur de Rephrase, Sam Angove, pour sa rapidité à répondre aux questions des utilisateurs.