

XML pour la gestion des actualités sur Internet : utilisation de RSS

Andrei Popescu-Belis
TIM / ETI, Université de Genève

Cours n°11

Qu'est-ce que RSS ?

- Une norme basée sur XML pour décrire le contenu des bulletins d'informations sur Internet (news)
 - très grande simplicité, grande polyvalence: codage générique des métadonnées pour des pages web
 - destinée à communiquer les nouveautés entre sites, et aux utilisateurs humains
- « Producteurs » de flux ↔ « Consommateurs » de flux
- Plusieurs versions et noms concurrents existent...
 - 1.0 = RDF Site Summary (ou parfois Rich Site Summary)
 - 0.9x, 2.0 = Really Simple Syndication

2

Définition des flux selon la documentation d'Internet Explorer

- Qu'est-ce qu'un flux?
 - Un flux, aussi appelé *flux RSS*, *flux XML*, *contenu syndiqué* ou *flux Web*, contient des informations fréquemment mises à jour et publiées par un site Web.
 - Un flux est généralement utilisé par des sites Web d'informations et de blog, mais il peut aussi être utilisé pour distribuer d'autres types de contenu numérique...
- Quels sont les différents formats de flux ?
 - Les formats les plus courants sont RSS et Atom. Les formats de flux sont constamment mis à jour vers de nouvelles versions.
 - Internet Explorer prend en charge les formats RSS 0.91, 1.0 et 2.0, ainsi que les formats ATOM 0.3 et 1.0.
 - Tous les formats de flux Web reposent sur le langage XML (Extensible Markup Language), un langage informatique à base de texte utilisé pour décrire et distribuer des données et des documents structurés.

3

Flux RSS : « producteurs »

- Producteur
 - un site Internet qui est mis à jour régulièrement, par exemple avec des dépêches concernant l'actualité (... notion très générale)
- Flux
 - document XML (format RSS) que le site met à jour régulièrement
 - énumère les « nouveautés » les plus récentes et contient des liens vers leur description complète
 - sorte de sommaire des nouveautés
 - ce document se trouve à une adresse particulière du site (URL)
 - appelée elle-même « flux » ou « canal RSS » ou « RSS feed »
 - un site peut mettre à disposition plusieurs flux
- Fonctionnement
 - il faut donc consulter régulièrement le flux RSS pour être tenu au courant des actualités → voir la description des « lecteurs RSS »

4

Contenu d'un fichier RSS

- Un fichier correspond à un canal (<channel>) ou flux
 - titre, URL (pour mémoire), description, langue
- Un canal contient un ou plusieurs « items » d'actualité
 - liste d'items qui change dans le temps
 - parfois toutes les minutes, selon le site
- Un item (<item>) correspond à une information nouvelle
 - titre, courte description, URL vers le contenu complet
 - le contenu complet n'est pas inclus dans le flux!
 - le flux n'offre que des liens vers des annonces/nouveautés
 - autres descripteurs d'un vocabulaire RDF
 - 3 vocabulaires officiels, dont DC + 19 autres proposés

5

Exemple de document RSS 0.91 (description des actualités d'un site à un moment t)

```
<rss version="0.91">
  <channel>
    <title>XML.com</title>
    <link>http://www.xml.com/</link>
    <description>XML.com features information for the XML community.</description>
    <language>en-us</language>

    <item>
      <title>Normalizing XML, Part 2</title>
      <link>http://www.xml.com/pub/a/2002/12/04/normalizing.html</link>
      <description>In this second and final look at applying relational normalization techniques to W3C XML Schema data modeling, Will Provost discusses when not to normalize, the scope of uniqueness and the fourth and fifth normal forms.</description>
    </item>

    <item>
      <title>SVG's Past and Promising Future</title>
      <link>http://www.xml.com/pub/a/2002/12/04/svg.html</link>
      <description>In this month's SVG column, Antoine Quint looks back at SVG's journey through 2002 and looks forward to 2003.</description>
    </item>
  </channel>
</rss>
```

6

Exemple de document RSS 1.0 (description des actualités d'un site à un moment t)

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns="http://purl.org/rss/1.0/"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <channel rdf:about="http://www.xml.com/cs/xml/query/q/19">
    <title>XML.com </title>
    <link>http://www.xml.com/</link>
    <description>XML.com features a rich mix of information.</description>
    <language>en-us</language>
  </channel>
  <items>
    <rdf:Seq>
      <rdf:li rdf:resource="http://www.xml.com/pub/a/2002/12/04/normalizing.html"/>
      <rdf:li rdf:resource="http://www.xml.com/pub/a/2002/12/04/svg.html"/>
    </rdf:Seq>
  </items>
  </channel>
  <item rdf:about="http://www.xml.com/pub/a/2002/12/04/normalizing.html">
    <title>Normalizing XML, Part 2</title>
    <link>http://www.xml.com/pub/a/2002/12/04/normalizing.html</link>
    <description>In this second and final look ...</description>
    <dc:creator>Will Provost</dc:creator>
    <dc:date>2002-12-04</dc:date>
  </item>
  <item rdf:about="http://www.xml.com/pub/a/2002/12/04/svg.html">
    <title>SVG's Past and Promising Future</title>
    <link>http://www.xml.com/pub/a/2002/12/04/svg.html</link>
    <description>In this month's SVG column, Antoine Quint ...</description>
    <dc:creator>Antoine Quint</dc:creator>
    <dc:date>2002-12-04</dc:date>
  </item>
</rdf:RDF>
```

7

Sites web fournissant des flux RSS

- Fournisseurs de contenu RSS → utilisent divers logos RSS



- la plupart des sites contenant des « actualités »
- journaux, dépêches d'agences, forums, blogs, etc.
- les sites se signalent eux-mêmes avec un logo
 - répertoires de flux: <http://www.Syndic8.com> ou <http://www.2rss.com>
 - et beaucoup d'autres → voir aussi les sites « RSS readers »

8

Flux RSS : « consommateurs »

- Fonctionnalités requises d'un « lecteur RSS »
 - afficher un/plusieurs canaux RSS (document XML) sous un format lisible, surtout les titres et les listes des nouveautés
 - offrir l'accès au contenu de chaque <item> en cliquant sur l'URL
 - retenir l'adresse de canaux RSS (marque-pages), aider à en découvrir de nouveaux
 - au fond, assez semblables aux navigateurs web
- Lecteurs RSS spécialisés : *RSS reader* ou *RSS aggregator*
 - agrégation d'informations (canaux), alerte, filtrage, recherche
- Utiles pour :
 - suivre les actualités (professionnelles ou non)
 - être informé de nouvelles entrées dans un blog, de nouveaux podcasts

9

Flux RSS : « consommateurs »

- Deux grands types (et beaucoup de variantes)
 - outils individuels sur chaque ordinateur (gratuits ou payants)
 - marque-pages dynamiques dans un navigateur web : FF, NS, IE-
 - modules à ajouter à un navigateur web :
Firefox → *Wizz RSS News Reader* ou *Sage*
 - programmes spécifiques
 - sites web spécialisés
 - page personnelle (mot de passe) combinant des canaux RSS
 - page générique contenant une zone « actualités » provenant de plusieurs flux
 - pour rechercher de tels outils
 - <http://allrss.com/rssreaders.html>

10

Marque-pages dynamiques (*live bookmarks*)

- Exemple
 - canal RSS « Nature/Science » de la BBC
 - adresse URL
http://newsrss.bbc.co.uk/rss/newsonline_uk_edition/sci/tech/rss.xml
- Démonstration
 - affichage de ce lien dans FF, IE, NS
 - ajout de la page comme marque-page dynamique
 - consultation ultérieure

11

RSS Readers

- Sites web spécialisés
 - composer sa propre actualité en cherchant parmi les canaux offerts
 - sauvegarder sa requête pour la consulter régulièrement
 - nécessitent une inscription: nom d'utilisateur et mot de passe (pour sauvegarder le profil)
 - exemples : Bloglines.com, MyYahoo, Netvibes, Google Reader, NewsGator, Rojo.com, NewsIsFree
- Logiciels permettant de sélectionner et lire les flux
 - permettent de consulter les canaux sur son propre ordinateur
 - exemples: Amphetadesk, FeedDemon, Awasu, Newsfire (MacOS)
- Modules de sites plus complexes
 - exemple du site Moodle de notre cours !

12

Comme afficher le contenu d'un flux RSS dans une page web ?

- Beaucoup de sites les utilisent
 - rendre le site plus attractif, plus évolutif, plus actuel
 - le choix des informations reflète la personnalité du site
 - nécessite des outils au niveau du serveur
- Difficulté : rendre le flux RSS/XML de lisible
= « parser » le fichier RDF/XML, puis utiliser une feuille de style...
- Solutions pour vos pages
 - voir <http://www.rss-specifications.com/display-rss.htm>
 - faire appel à un convertisseur externe grâce à quelques lignes de Javascript dans votre page : comment les écrire ?
 - par exemple utiliser un *générateur automatique de code* disponible à :
 - <http://www.blinkbits.com/feed/build.php>
 - <http://www.rss-to-javascript.com/p/138.html>
 - inconvénient : appel nécessaire à un site extérieur pour effectuer la conversion → démonstration avec « [essai.html](#) »

13

Exercices (1)

- Cherchez des sites qui fournissent des canaux RSS
 - repérez le logo RSS/XML sur le site
 - soit selon votre inspiration (p.ex. vos journaux favoris)
 - soit avec <http://www.Syndic8.com>
- Choisissez un canal RSS et ouvrez-le avec Firefox
 - ajoutez-le ensuite à vos favoris
 - inspectez avec Ctrl+B la barre latérale des marque-pages
- Essayez aussi avec Internet Explorer
- Au choix
 - A. Installer et utiliser Wizz RSS News Reader
 - B. Aller au transparent suivant pour afficher un flux sur votre page web

14

Exercices (2) : pour aller plus loin

- Essayez d'ajouter à votre page Internet (HTML) un paragraphe ou un cadre avec des nouvelles provenant d'un canal RSS de votre choix
 - choisissez un canal RSS et notez son adresse précise
 - allez à <http://www.blinkbits.com/feed/build.php> et rentrez l'adresse du canal et vos options d'affichage
 - générez le code Javascript en cliquant sur « *generate* »
 - copiez ce code et collez-le dans votre page HTML
 - affichez votre page : comment s'affiche le canal ?

15