

Applications d'XML : RSS – présentation & exercices

Andrei Popescu-Belis
TIM / ETI, Université de Genève

Cours n°11 — 2.6.2005

Présentation de RSS

RDF Site Summary

Qu'est-ce que RSS ?

- Une norme (basée sur XML, bien sûr) 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
- Plusieurs versions concurrentes existent...
 - 1.0 = RDF Site Summary
 - 0.9x, 2.0 = Really Simple Syndication

3

Principe (1)

- Fournisseurs d'informations / actualités
 - sites faisant l'objet de mises à jour fréquentes
 - diffusion de « news » au format RSS/XML
 - à une adresse particulière du site (URL) = nom du fichier RSS
 - *RSS feed* / canal ou flux ou fil RSS
- Utilisateurs des informations
 - outils spécialisés
 - combinent plusieurs canaux RSS
 - objectifs : agrégation d'informations, alerte, filtrage, recherche
 - « RSS reader » ou « RSS aggregator »

4

Principe (2)

- Les lecteurs RSS
 - interprètent le canal RSS pour l'afficher (graphiquement)
 - offrent ensuite accès à la source
- Trois types
 - sites spécialisés combinant des canaux RSS
 - paramétrables pour les utilisateurs enregistrés
 - outils individuels dédiés
 - pages Internet contenant une zone « actualités »
- Fonctionnement
 - les outils consultent régulièrement (p.ex. toutes les heures) les sites qui fournissent les flux RSS afin de mettre à jour les actualités affichées

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>

    <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

Contenu d'un fichier RSS

- un fichier = un canal (<channel>)
 - titre, URL (lien), description, langue
- un canal = un ou plusieurs « items » d'actualité → liste
 - change dans le temps (parfois toutes les minutes)
- un item = une information nouvelle (<item>)
 - titre, URL (lien), description
 - autres descripteurs d'un vocabulaire RDF
 - 3 vocabulaires officiels, dont DC + 19 autres proposés

8

Outils et sites RSS

- Fournisseurs de contenu RSS → divers logos RSS



- la plupart des sites contenant des « actualités »
 - journaux, dépêches d'agence, forums, etc.
 - voir liste à www.Syndic8.com (« syndicate »)
- Outils pour lire des canaux RSS
 - cf. les trois types mentionnés ci-dessus
 - souvent gratuits
 - pratiques pour suivre les actualités professionnelles ou non
 - chercher « RSS News Readers » à <http://directory.google.com>

9

RSS Readers : sites Internet

- Principe
 - on compose sa propre actualité en cherchant parmi les canaux offerts
 - on peut sauvegarder sa requête pour la consulter régulièrement (ou être alerté)
 - nécessitent une inscription (pour sauver son profil)
- Exemples
 - Bloglines, My Yahoo!
 - NewsGator
 - Meerkat, NewsIsFree

10

RSS Readers : outils individuels

- Logiciels qui permettent de sélectionner et lire les flux
 - gratuits ou payants
 - permettent de consulter les canaux et les sauver sur son propre ordinateur
- Exemples
 - Windows : Amphetadesk, Newz Crawler, FeedDemon, Awasu,
 - Mac OS X : Newsfire, NetNewsWire
- Alternative : « plugin » dans un navigateur Internet
 - **Sage pour Firefox** : <http://sage.mozdev.org/>
 - très simple à installer et à utiliser
 - utilise les icônes de Firefox

11

RSS dans les pages Internet

- Beaucoup de sites les utilisent
 - rendre le site plus attractif
 - le choix des infos reflète la personnalité du site
 - nécessite des outils au niveau du serveur
- Difficulté : transformer le flux RSS/XML en quelque chose de lisible
 - « parser » le fichier RDF/XML, puis utiliser une feuille de style...
- Solution simple (recommandable)
 - insérer quelques lignes de Javascript dans votre page
 - utiliser le *générateur automatique de code* disponible à:
 - <http://jade.mcli.dist.maricopa.edu/feed/>
 - <http://www.globalsyndication.com/resources>
 - seul inconvénient : fait appel à un site extérieur à votre page
 - voir démonstration

12

Conclusion

- XML est un support pour de nombreux standards dans le monde du *semantic web*
- L'utilisation grand public de ces standards, comme dans le cas de RSS, ne requiert pas la maîtrise totale de XML... mais juste la connaissance des principes
- L'exploitation d'un standard fait appel à des notions plus avancées d'XML
 - p.ex., pour produire soi-même un canal RSS associé à un site...
- Importance de l'apprentissage par l'exemple

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. journaux)
 - soit avec www.Syndic8.com
- Réalisez des recherches d'actualités sur un sujet donné avec [Meerkat](#)
- Installez l'outil Sage pour Firefox
 - <http://sage.mozdev.org/>
 - choisissez un canal RSS et ajoutez-le à vos favoris avec Sage et l'icône de Firefox

14

Exercices (2)

- Essayez de créer une page Internet (HTML) avec des nouvelles provenant d'un canal
 - écrivez une page très simple
 - choisissez un canal RSS et copiez son adresse précise
 - allez à <http://jade.mcli.dist.maricopa.edu/feed/> ou <http://www.globalsyndication.com/resources> 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 locale
 - affichez votre page : comment s'affiche le canal ?