

Standards pour la description sémantique et les métadonnées : RDF et Dublin Core

Andrei Popescu-Belis
TIM / ETI, Université de Genève

Cours n°10b

Resource Description Framework (RDF) : <http://www.w3.org/RDF>

- Un cadre pour décrire les « ressources » disponibles sur Internet
- Norme unifiée pour écrire les méta-données
 - d'une page Internet
 - d'un catalogue
 - d'un journal en ligne
 - etc.
- Possède une description abstraite ainsi qu'une syntaxe XML

2

Principes (1)

- Les informations apparaissent comme une série de « propositions »
 - sujet (= info à propos de quoi ?)
 - prédicat (= que dit-on à propos du sujet ?)
 - objet (= argument du prédicat)
- Exemple
 - « personne_X a_pour_téléphone numéro_Y »
- Peuvent être codées en XML

3

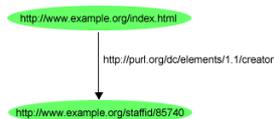
Principes (2)

- Les « sujets » ou ressources sont des URIs
 - Uniform Resource Identifiers
 - analogues aux adresses Internet
 - sans indiquer forcément un contenu
 - ce sont juste des indicateurs uniques
 - les URL (Uniform Resource Locators) sont des URI, mais tous les URI ne sont pas des URL
- Les prédicats appartiennent à un domaine d'application spécifié par un URI
- Les objets sont soit également des ressources, soit des données sous forme de nombre, chaîne, etc.

4

Exemple

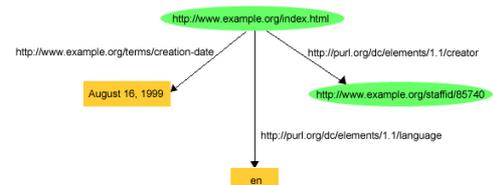
- Encodage en RDF de la description:
« La page <http://www.example.org/index.html> a été créée par l'utilisateur n°85749 »



```
<rdf:Description rdf:about="http://www.example.org/index.html">
  <dc:creator>85740</dc:creator>
</rdf:Description>
```

5

Exemple plus long (1)



6

Exemple plus long (2)

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:exterms="http://www.example.org/terms/">
  <rdf:Description rdf:about="http://www.example.org/index.html">
    <exterms:creation-date>August 16, 1999</exterms:creation-date>
    <dc:language>en</dc:language>
    <dc:creator rdf:resource="http://www.example.org/staffid/85740"/>
  </rdf:Description>
</rdf:RDF>
```

7

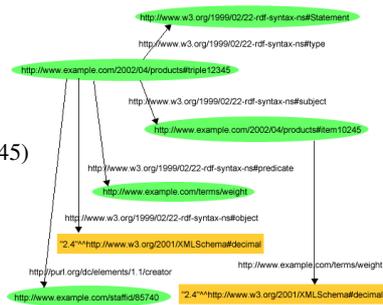
Possibilités de RDF

- RDF est un modèle abstrait mais peut être encodé en XML
→ on parle de RDF/XML
- Le système sujet-prédicat-objet, avec des ressources (URI) et des valeurs simples (chaînes, nombres, etc.) est très expressif
 - décrire les propriétés *sémantiques* de n'importe quelle ressource
- Autres fonctionnalités
 - création de « sujets composés » (plusieurs ressources, avec ordre ou non)
 - création de « sujets blancs » (réfèrents sans noms)
 - combinaison de plusieurs conventions sémantiques
 - divers raccourcis et abréviations

8

Exemple encore plus complexe

John Smith
(id n° 85740)
a dit que
la tente
(article n° 10245)
pesait 2,4 kg



9

Définir un vocabulaire RDF

- Par des personnes ou organisations
- Bases de RDF
 - graphes avec des nœuds et des arcs = descriptions
 - RDF spécifie de façon formelle leur syntaxe et leur "sémantique" aussi
- Contenu d'un vocabulaire
 - les types de sujets/objets autorisés
 - les prédicats
 - les contraintes sur les arguments des prédicats (quels types vont avec quels prédicats)
- Le mécanisme pour le faire est défini dans le document *RDF Vocabulary Description Language 1.0: RDF Schema*

10

Conclusion sur RDF

- RDF permet la définition d'une sémantique formelle des ressources, sous la forme d'une série de *déclarations*
 - préciser le sens de chaque « prédicat »
 - préciser les valeurs possibles pour les variables (chaînes de caractères, nombres, dates, etc.) et les contraintes
- *Note* : les idées de RDF plongent leurs racines dans l'intelligence artificielle et la représentation des connaissances, notamment la théorie des graphes conceptuels proposée par John Sowa, les représentations logiques, les BD relationnelles
- Exemples de vocabulaires
 - **DC**: **Dublin Core** Metadata Initiative (DCMI parfois)
 - **RSS** : RDF Site Summary

11

Dublin Core

- Standard de description des ressources inspiré de l'univers des bibliothèques
- Chaque « ressource » (livre, page Internet, etc.) est décrite par
 - quinze descripteurs principaux
 - tous optionnels et répétables
 - des spécificateurs supplémentaires pour chaque descripteur
- DC est un vocabulaire assez répandu
 - origine du nom : conférence à Dublin, Ohio, USA

12

Dublin Core: les 15 descripteurs

- **Title:** A name given to the resource.
- **Creator:** An entity primarily responsible for making the content of the resource.
- **Subject:** The topic of the content of the resource.
- **Description:** An account of the content of the resource.
- **Publisher:** An entity responsible for making the resource available
- **Contributor:** An entity responsible for making contributions to the content of the resource.
- **Date:** A date associated with an event in the life cycle of the resource.
- **Type:** The nature or genre of the content of the resource.
- **Format:** The physical or digital manifestation of the resource.
- **Identifier:** An unambiguous reference to the resource within a given context.
- **Source:** A reference to a resource from which the present resource is derived.
- **Language:** A language of the intellectual content of the resource.
- **Relation:** A reference to a related resource.
- **Coverage:** The extent or scope of the content of the resource.
- **Rights:** Information about rights held in and over the resource.

13

Exemple de description RDF/XML avec les descripteurs DC

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/"
  <rdf:Description rdf:about="http://www.dlib.org">
  <dc:title>D-Lib Program - Research in Digital Libraries</dc:title>
  <dc:description>The D-Lib program supports the community of people
  with research interests in digital libraries and electronic
  publishing.</dc:description>
  <dc:publisher>Corporation For Research Initiatives</dc:publisher>
  <dc:date>1995-01-07</dc:date>
  <dc:subject>
  <rdf:Bag>
  <rdf:li>Research; statistical methods</rdf:li>
  <rdf:li>Education, research, related topics</rdf:li>
  <rdf:li>Library use Studies</rdf:li>
  </rdf:Bag>
  </dc:subject>
  <dc:type>World Wide Web Home Page</dc:type>
  <dc:format>text/html</dc:format>
  <dc:language>en</dc:language>
  </rdf:Description>
</rdf:RDF>
```

14

Conclusion sur Dublin Core

- Modèle de description des méta-données
 - peut être implémenté en XML de plusieurs façons
 - implémentation la plus logique : RDF/XML
- Utilisé par plusieurs initiatives visant à créer des catalogues communs d'archives
 - Open Archives Initiative (OAI)
 - <http://www.openarchives.org>
 - Open Language Archives Community (OLAC)
 - <http://www.language-archives.org/>

15

Exercices

- Objectif = créer un fichier de description RDF (ou éventuellement DC) pour votre page d'accueil
 - cherchez des fichiers RDF sur Internet pour vous en inspirer
 - regardez aussi ces modèles
 - regarder le lien tout en bas de la page <http://www.semanticweb.org/> (lien "rdf/xml as text/xml")
 - mieux, regarder le site <http://dublincore.org/>
 - au bas de chaque page, il y a un lien vers les métadonnées associées (faire « afficher source » pour voir les balises)
 - par exemple <http://dublincore.org/index.shtmlrdf>
 - sur votre site étudiant, ajoutez un document `index.html.rdf` qui décrit en RDF votre page d'accueil (vérifiez qu'il est bien formé et valide)

16