Introduction au Web sémantique

ressources

site du W3C w3c.org websemantique.org

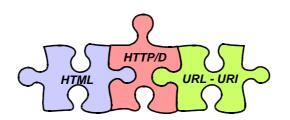
Le wiki de l'équipe-projet Edelweiss à Sophia-Antipolis www-sop.inria.fr/edelweiss/

1

Le Web : une couche applicative au dessus d'Internet

- 1990 HTML, HTTP (Tim Berners Lee au CERN)
- 1991 navigateur sous NeXTStep
- 1992 HTML 1.0
- 1994 W3C (World Wide Web Consortium)
- 1994 URL (Uniform Resource Locator)
- 1995 HTML 2.0
- 1996 HTTP 1.0
- La suite : tableaux, texte autour images, applets, indices/exposants, scripts, stylesheets, ...

Naissance du Web



le puzzle du web ...

2

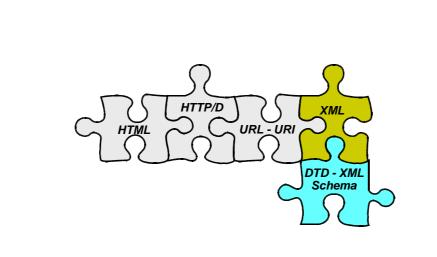
- 1998 XML 1.0 (Extensible Markup Language)
 - représenter et non présenter (séparation entre contenu et présentation)
 - un format textuel d'échange de données structurées
 - standard pour définir des langages balisés
- Web structuré accessible comme Base de Données
- Famille de langages : MathML, CML, SVG, XMI, XHTML, XFRAMES, ...

XML 4

- DTD (Document Type Definition)
 - une DTD définit les balises autorisées, leurs attributs et leur enchaînement
- 2004 XML Schéma
 - contraintes sur structure (namespaces, éléments, attributs, cardinalité) et contenu (datatypes et types prédéfinis, entités, notations, expressions régulières)
 - notion de type et héritage pour définitions éléments, attributs, et datatypes (extension & restriction)
 - valeurs par défaut des éléments et des attributs
 - Spécification partielle (élément / attribut quelconque)

DTD, XML Schema

5



... le puzzle du web ...

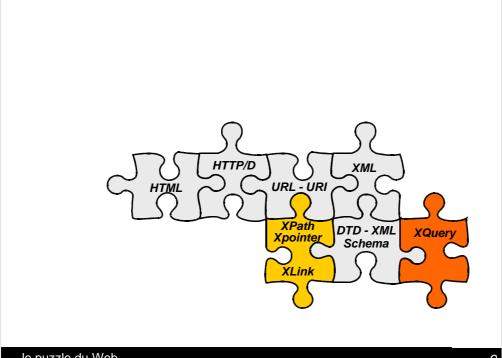
- 1999 XPath 1.0 (XML Path Language)
 - description des chemins dans un document XML
- 2001 Xlink 1.0 (XML Linking Language)
 - la généralisation XML du concept de lien du HTML
 - plus expressif (destinations multiples, contrôle des déclencheurs, adaptation du comportement, ...)
- 2003 XPointer 1.0 (XML Pointer Language)
 - extension des URL pour pointer sur des éléments d'un document XML même si le fragment est sans ID
 - fait une connexion entre XPath et les URL

Xpath, Xlink, Xpointer

7

- 2006 XQuery (XML Query Language)
 - langage de requête sur les structures XML, inspiré de SQL
 - S'appuie sur les systèmes d'adressage XPath, XLink, XPointer

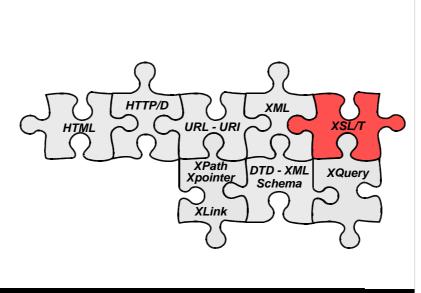
XQuery



... le puzzle du Web ...

- 2001 XSL 1.0 (XML Stylesheet Language)
 - feuilles de style pour XML
- 1999 XSLT 1.0 (XSL Transformation)
 - transformation récursive d'un arbre XML source en un arbre XML cible: tester et filtrer, modifier des valeurs, ajouter des éléments, réorganiser (trier), etc.

XSLT



... le puzzle du Web ...

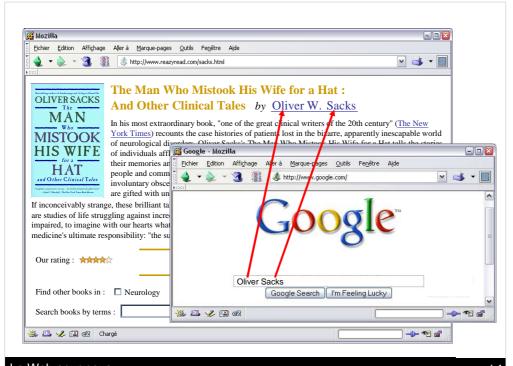
11

- Le Web est donc
 - un ensemble de ressources (données et pages)
 - bien outillé pour
 - la représentation de données structurées
 - leur restitution sous forme de page Web
 - l'adressage des ressources
 - la recherche de ressources
 - La recherche d'éléments à l'intérieur des ressources
- Mais les ressources stockées et manipulées dans le cadre du Web sont purement syntaxiques

Du Web au Web sémantique.

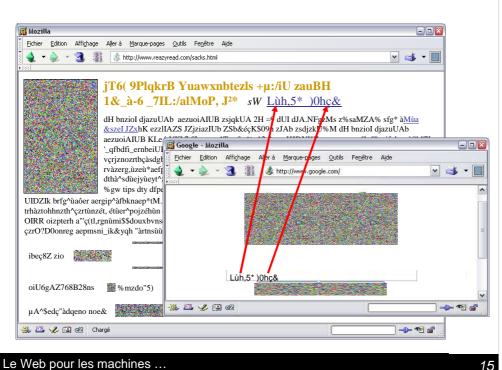
<u> 12</u>





Le Web pour nous ...

<u> 14</u>



- Dans le Web classique, pas d'interprétation
 - manipulation formelle, syntaxique des données
- Dans les SI « traditionnels », la sémantique n'est pas forcément nécessaire à l'interaction homme-machine
 - les informations relèvent d'un domaine borné et souvent technique
 - pas ou peu d'ambiguité
 - système fermé à destination de spécialistes
- Le Web est un système ouvert à tous, couvrant tous les domaines

- La masse des informations stockées sur le Web augmente à vitesse exponentielle
 - 5.10⁶ To (1 To = 10¹² octets) en 2005 (la Library of Congress stocke 20 To de données texte)
 - en 2005, le nombre de sites Web a augmenté de 17,5 millions
 - la production mondiale de données mémorisées double chaque année
- Le problème du Web n'est plus d'augmenter la taille des « autoroutes de l'information », mais de concevoir et réaliser des systèmes permettant de filtrer les informations et de les délivrer de façon « intelligente »

Du Web au Web sémantique

- « The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation »
 Tim Berners-Lee, James Hendler, Ora Lassila in The Semantic Web, Scientific American, May 2001
- Ajout d'une couche de connaissances au dessus du Web pour que :
 - les applications Web interprètent les données comme les utilisateurs afin d'améliorer l'interaction
 - l'interopérabilité et le dialogue entre applications soit possible

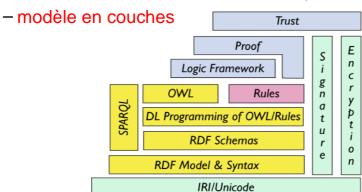
- La couche connaissance du Web sémantique contient
 - des annotations sémantiques (RDF)
 - des ressources sémantiques (ontologies en RDFS ou OWL)
 - des outils de manipulations des annotations et ontologies (SPARQL, ...)
- D'autres éléments sont en cours d'ajout
 - Des règles pour augmenter l'expressivité des ontologies OWL

– ...

La sémantique dans le Web sémantique

19

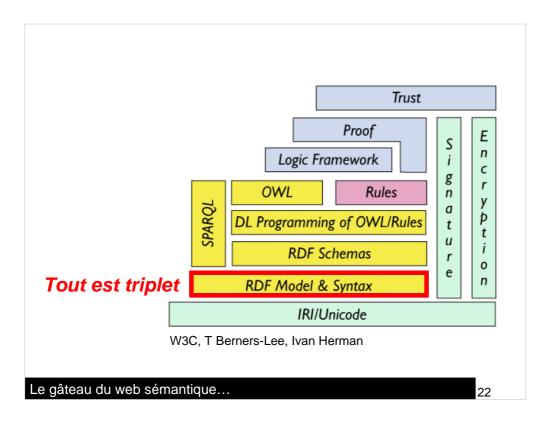
- Les 4 principaux standards du Web sémantique
 - RDF: un modèle de triplets pour décrire et connecter des ressources anonymes ou identifiées par un URI
 - SPARQL: un langage de requête sur les graphes RDF
 - RDFS est un langage de descriptions légères
 - OWL: 3 couches d'extension de l'expressivité (logique)



Le modèle en couches du Web sémantique

Resource Description Framework





• 1999 – RDF (Ressource Description Framework)



- Annotation sémantique des ressources
- Assertion de liens entre ressources (donner du sens, c'est lier des informations ou connaissances)
- Plus riche qu'une annotation syntaxique par mot-clé
- Triplet RDF <entité, propriété, valeur>
 - Décrit une entité (identifié par une URI)
 - Associe au sujet une propriété (identifiée par une URI)
 - Donne une valeur à la propriété. Une valeur est soit une ressource identifiée par une URI, soit une valeur primitive

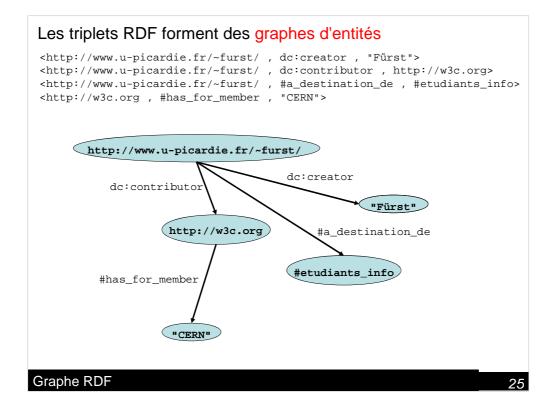
Triplet RDF 23

• Exemples :

```
<http://www.u-picardie.fr/~furst/ , dc:creator , "Fürst">
<http://www.u-picardie.fr/~furst/ , dc:contributor , http://w3c.org>
<http://www.u-picardie.fr/~furst/ , #a_destination_de , #etudiants_info>
<http://w3c.org , #has_for_member , "CERN">
```

- Les URL introduites par # sont définies localement
- Des prédicats sont définis dans des vocabulaires existants :
 - dc (Dublin Core) : schéma de métadonnées pour les documents (dublincore.org)
 - Il existe d'autre vocabulaires (ex. foaf Friend of a Friend, www.foaf-project.org)
 - rdf définit aussi ses propres prédicats

Triplet RDF 24



- Syntaxe XML pour RDF: RDF possède une syntaxe XML (mais ne s'y réduit pas!)
- Un document RDF est une liste de descriptions inclues dans des balises rdf:RDF
- Une description est introduite par la balise rdf:Description et porte sur une entité désignée par une URI introduite par l'attribut rdf:about
- A la place de l'attribut rdf:about, on peut utiliser rdf:ID pour donner une URL relative

Vocabulaire RDF 26

```
<http://www.u-picardie.fr/~furst/ , dc:creator , "Fürst">
<http://www.u-picardie.fr/~furst/ , dc:contributor , http://w3c.org>
<http://www.u-picardie.fr/~furst/ , #a_destination_de , #etudiants_info>
<http://w3c.org , #has_for_member , "CERN">
    <?xml version="1.0"?>
    <rdf:RDF xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
             xmlns:dc = "http://dublincore.org/2008/01/14/dcelements.rdf#"
             xmlns:ex = "http://www.u-picardie.fr/~furst/exemples.rdf#">
        <rdf:Description rdf:about="http://www.u-picardie.fr/~furst/">
            <dc:creator>Fürst</dc:creator>
            <dc:contributor>
                 <rdf:Description rdf:about="http://w3c.org">
                     <ex:hasformember>CERN</ex:hasformember>
                </rdf:Description>
            </dc:contributor>
            <ex:a_destination_de rdf:resource="ex:etudiants_info"/>
        </rdf:Description>
    </rdf:RDF>
```

Syntaxe XML pour RDF

27

Contenu du fichier http://www.u-picardie.fr/~furst/exemples.rdf

Syntaxe XML pour RDF

<u> 28</u>

- Utilisation de l'attribut rdf:ID
 - rdf:about="truc#blabla" est équivalent à rdf:ID="blabla" si l'URI de référence est truc

Vocabulaire RDF

29

Vocabulaire XML propre à RDF

Mots réservés

RDF, Description, ID, about, parseType, resource, li, nodeID, datatype

Noms de classes

Seq, Bag, Alt, Statement, Property,
XMLLiteral, List

Noms de propriétés

subject, predicate, object, type, value,
first, rest, _n (où n est un entier plus grand que 0 et ne
commençant pas par 0)

Noms de ressources

nil

Vocabulaire RDF

- Les ressources sont identifiables par des URI
 - Si l'URI est un URL alors ressource du Web
 - Si non, URI d'une ressource physique, abstraite, etc.
 - Jointure entre les assertions même si elles sont distribuées
 - Liant entre les couches réseau (Internet, Web, Web sémantique)
- RDF est un modèle de métadonnées ouvert
 - Vocabulaire extensible basé sur les URI et XML schema datatypes
 - Autorise quiconque à faire des déclarations sur n'importe quelle ressource

Modèle de triplets

31

- Il existe un autre langage pour exprimer plus simplement des triplets RDF: le langage N3 développé par T. Berners Lee
 - Syntaxe non XML
 - Vise une manipulation humaine

```
@prefix dc: <http://purl.org/dc/elements/1.1/>.
    <http://www.u-picardie.fr/~furst/>
        dc:title "Cours d'informatique";
        dc:creator "Fürst".
```

Langage N3 32

Chaque triplet représente un <u>prédicat binaire</u> en logique

```
<http://www.u-picardie.fr/~furst/ , dc:creator , "Fürst">
<http://www.u-picardie.fr/~furst/ , dc:contributor , http://w3c.org>
<http://www.u-picardie.fr/~furst/ , #a_destination_de , #etudiants_info>

dc:creator(http://www.u-picardie.fr/~furst/, "Fürst")
=> dc:contributor(http://www.u-picardie.fr/~furst/, http://w3c.org)
#a_destination_de(http://www.u-picardie.fr/~furst/, #etudiants_info)
```

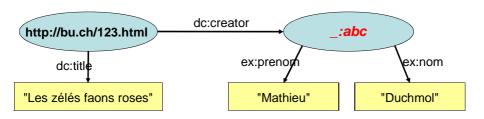
- <u>Sémantique formelle</u>: RDF est un sous-ensemble de la logique du premier ordre
 - Avec : prédicats binaires, quantification existentielle, conjonction
 - Sans : disjonction, négation, quantification universelle
- Tout énoncé RDF est considéré comme vrai et RDF est monotone i.e. ce qui est vrai et ce que l'on peut déduire reste vrai si l'on rajoute de nouveaux énoncés.

Modèle logique

33

•Une ressource peut ne pas être identifiée

•Quantification existentielle : ∃ x dc:author(http://bu.ch/l23.html,x) ∧ ex:nom(x,"Duchmol") ∧ ex:prenom(x,"Mathieu")



Les nœuds anonymes (blank nodes)

- En cas de noeud anonyme, les parseurs RDF créent des ID
- On peut également introduire un ID local pour identifier les noeuds anonymes (vital pour les sérialisations)

Les nœuds anonymes avec ID

35

• Il est possible d'omettre les balises rdf:Description pour un noeud anonyme, en utilisant l'attribut rdf:parseType

 Autre utilisation de rdf:parseType : obliger le parseur à ignorer la structure du contenu

Utilisation de parseType

Exercice

- Récupérer le fichier human.rdf situé sur la page http://www.u-picardie.fr/~furst/onto.html
- Question 1 : Quel est le namespace utilisé pour les instances / ressources crées dans ce fichier? Par quel mécanisme l'association entre instances et namespace est-elle faite?
- Question 2 : Quel est le namespace du schéma RDF utilisé et comment est-il associé aux balises?
- Question 3: Trouver toutes les informations sur John.

Exercice

37

- Réification d'un triplet : rendre un triplet explicite pour pouvoir en parler i.e. l'utiliser comme le sujet ou l'objet d'une propriété
 - Un triplet est réifié par un statement (balise rdf:Statement)
 - Le statement fait du triplet une ressource
 - Cette ressource peut être décrite à son tour

Réification d'un statement

- · XML schema datatypes
 - Les littéraux standards sont des chaînes de caractères
 - Pour typer les valeurs littérales, RDF repose sur les datatypes de XML Schema

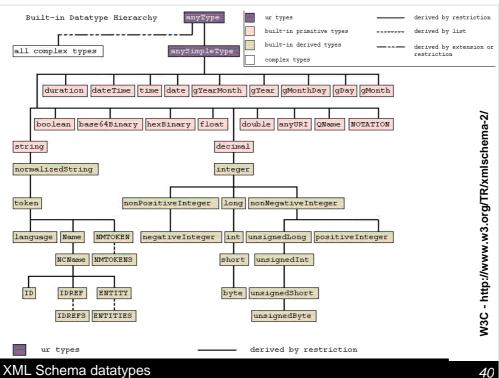
```
<rdf:Description>
    <ns:age rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">12</ns:age>
</rdf:Description>
```

Notation en N3

ns:id1 ns:age "22"^^xsd:integer

Datatypes

39



XML Schema datatypes

- On peut typer les ressources par l'attribut rdf:type
 - Le type peut être une rdf:Resource, une rdf:Property ou un rdf:Statement
 - Une ressource peut avoir plusieurs types

· Autre syntaxe possible :

```
<ex:Personne rdf:about="urn://~furst">
...
</ex:Personne>
```

• L'instanciation d'un type suffit à faire exister une ressource

```
_:x rdf:type http://www.ugb.sn/schema#Personne
```

Typage de ressource

11

• On peut spécifier une langue avec xml:lang

```
<ex:Personne rdf:about="urn://~furst">
    <ex:job xml:lang="fr">Maitre de Conférences</ex:job>
    <ex:job xml:lang="en">Associate professor</ex:job>
</ex:Personne>
```

• En N3:

```
<urn://~furst>
   ex:job "Maitre de Conférences"@fr
   ex:job "Associate professor"@en
```

 <u>Attention</u>: les littéraux avec langue et sans langue sont différents

```
"Truc" # "Truc"@en # "Truc"@fr
```

Langue 42

 Un triplet peut avoir pour objet un groupe non ordonné de ressources ou de littéraux

 L'accès aux valeurs se fait par l'attribut rdf:_n (n entier) qui peut également être utilisé dans la déclaration

Groupe simple sans ordre ni sens

12

 Un triplet peut avoir pour objet un groupe ordonné de ressources ou de littéraux

L'accès aux valeurs se fait de la même manière

Séquence 44

 Un triplet peut avoir pour objet un groupe de ressources ou de littéraux alternatifs (le premier est le plus usité)

Alternatives 45

 Une ressource peut être décrite comme liste exhaustive et ordonnée de constituants

```
<rdf:Description rdf:about="#Semaine">
    <ex:jours rdf:parseType="Collection">
        <rdf:Description rdf:about="#Lundi"/>
        <rdf:Description rdf:about="#Mardi"/>
        <rdf:Description rdf:about="#Mercredi"/>
        <rdf:Description rdf:about="#Jeudi"/>
        <rdf:Description rdf:about="#Vendredi"/>
        <rdf:Description rdf:about="#Samedi"/>
        <rdf:Description rdf:about="#Dimanche"/>
        </ex:jours>
    </rdf:Description>
```

• Les propriétés rdf:first et rdf:rest permettent de parcourir une collection

Collection 46

Description équivalente :

Collection 47

- RDF : standard permettant l'échange de métadonnées sur le web et leur traitement automatique
- Modèle & syntaxe d'annotation :
 - Modèle simple avec une sémantique formelle (Graphes RDF au même niveau que l'arbre XML)
 - Format d'échange basé sur une syntaxe XML (RDF/XML)
- Le sens donné aux annotations provient des liens entre ressources (et entre métadonnées par réification)

Resource Description Framework

<u>48</u>

Limitations de RDF

Pas de restriction sur les objets et sujets d'une propriété

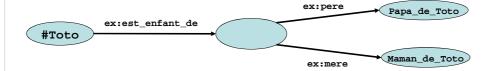
 Pas de lien de généralisation/spécialisation entre types ou entre propriétés (rdf:type lie uniquement une instance et un type)

 Expressivité limitée : pas de quantification universelle, pas de négation, pas de disjonction

Limitations de RDF

49

- Représenter les relations n-aires avec plusieurs objets
 - Choisir un sujet principal
 - Réifier la relation par une ressource anonyme
 - Lier la relation réifiée aux différents objets



Relations n-aires

- Parseurs RDF existants :
 - http://www.w3.org/RDF/Validator/
 - Jena (HP lab mais opensource) : http://jena.sourceforge.net/
 - Sesame (opensource) : http://www.openrdf.org/
 - ICS-Forth RDF suite
 - CWM
 - CORESE (INRIA)
 - Redland RDF Framework (en C)

–

Parseurs 51

Exercice

 Question 1 : en utilisant le vocabulaire défini sur la page http://xmlns.com/foaf/0.1/, écrire un fichier RDF exprimant les informations suivantes : Robert Duchmol est une personne surnommée Bob qui connait Marie Tartempion qui est membre de l'association des Amis de RDF.

Valider votre fichier sur http://www.w3.org/RDF/Validator/

 Question 2 : sans modifier les déclarations précédentes, ajouter l'information suivante : Bob a pour fratrie, dans l'ordre, Justine et Marcel

Valider votre fichier sur http://www.w3.org/RDF/Validator/