

Le web sémantique

Présentation du 6 mai 2017 à LinuQ

François Pelletier

Édité le 3 mai 2017



Le World Wide Web

Web Sémantique

Composantes du web sémantique

Projets du web sémantique

Le World Wide Web

- ▶ Une immense collection d'information et de données qui peuvent être accédées via Internet en utilisant le protocole HTTP. On utilise principalement un navigateur (Firefox, Chromium, Konqueror).
- ▶ Introduit au CERN par Tim Berners-Lee.
- ▶ Documents hypertextes connectés par des hyperliens.

Le premier navigateur était WorldWideWeb for NeXT.

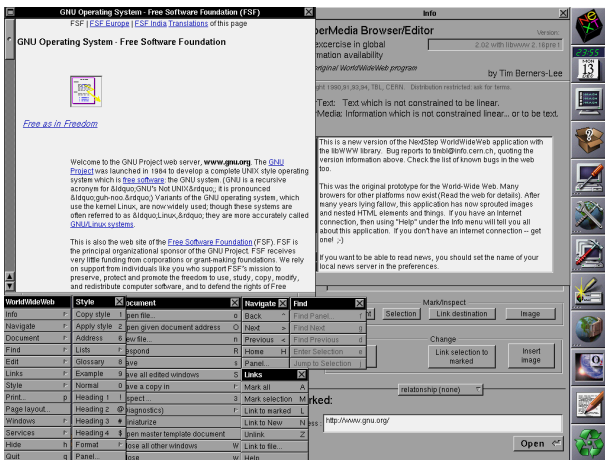


Figure – WorldWideWeb for NeXT

Qui standardise le web ?

- ▶ **IETF** : Internet Engineering Task Force
Définit les protocoles de communication
- ▶ **W3C** : World Wide Web Consortium
Définit les standards du WWW : HTML, CSS, XML, SVG, OWL, ...



Figure – Un seul web (2004)

Les Uniform Resource Identifier servent à identifier les différentes ressources du WWW. Il y en a deux types principaux :

- ▶ URN : Uniform Resource Name : Basée sur le nom de la ressource
- ▶ URL : Uniform Resource Locator : Basée sur l'adresse d'information de la ressource.
- ▶ Caractéristiques :
 - ▶ Universalité : peut être utilisée peu importe le service
 - ▶ Unicité : Chaque ressource est identifiable uniquement
 - ▶ Extensibilité : Permet d'identifier de nouvelles ressources
 - ▶ Fixabilité : Une URI soit être lisible par l'humain et imprimable, pas seulement électronique.

Les époques du web

- ▶ Web 1.0 : Le web statique
 - ▶ Space Jam (lien)
 - ▶ jayj.dk
- ▶ Web 2.0 : Création de contenu et participation
 - ▶ LiveJournal
 - ▶ MySpace
 - ▶ Facebook
- ▶ Web 3.0 : Web sémantique et d'exécution. Données ouvertes, services web et interfaces de programmation (API)
 - ▶ DBPedia
 - ▶ ProgrammableWeb

Pourquoi un web sémantique ?

Le web actuel a ses limites :

- ▶ Des milliards de documents sur le web, seule une petite partie est indexée par les moteurs de recherche.
- ▶ Le nombre double à tous les 6 mois
- ▶ Il devient difficile de :
 - ▶ trouver de l'information pertinente
 - ▶ différencier les données et la publicité
 - ▶ associer de la crédibilité à une information
- ▶ Il y a beaucoup de redondance, de copies et ça rend le web de mois en mois efficace.
- ▶ Un robot ne peut distinguer la signification, le contexte et la pertinence de l'information.
- ▶ Très peu de métadonnées sur la signification des documents et de leurs parties. Toutes les métadonnées doivent être créées par l'auteur.

Problèmes à résoudre

Deux types de problèmes abordés par le web sémantique :

- ▶ La recherche d'information : la recherche par mots clés apporte beaucoup de résultats peu pertinents. Besoin d'un système de terminologie formel : les ontologies. Hiérarchie de concepts et langage commun.
- ▶ L'extraction d'information : seul les humains peuvent extraire l'information correctement. Les robots manquent de contexte et de connaissances générales du monde.

Exemples d'usages potentiels

- ▶ Recherche contextuelle et personnalisée : Utilise les attributs et les connaissances de celui qui fait la recherche pour éliminer les résultats peu pertinents
- ▶ Planification sous contraintes : On effectue une planification où on a des contraintes de ressources, budgétaires, ...

Objectifs du web sémantique

- ▶ Organisation des connaissances en domaines séparés selon leur signification
- ▶ Outils automatique de maintenance, de correction des incohérences et d'extraction de nouvelles connaissances
- ▶ Remplacer la recherche par mots-clés par la recherche basée sur le contenu par une série de questions et réponses.
- ▶ Recherche orientée vers l'expérience utilisateur
- ▶ Extraction et présentation de connaissances
- ▶ Répondre à des questions à partir du contenu de plusieurs documents
- ▶ Contrôler l'accès à différents niveaux d'information

Le gâteau du web sémantique

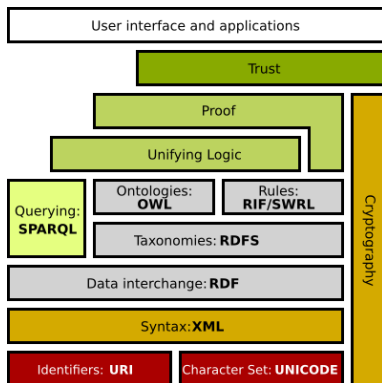


Figure – Le gâteau du web sémantique

Composantes du web sémantique : le XML

- ▶ XML signifie eXtensible Markup Language.
- ▶ Il sert à stocker et à transporter des données. Il est possible d'inclure des métadonnées via les étiquettes.
- ▶ Il est à la fois lisible par l'humain et par l'ordinateur.
- ▶ Il est donc approprié pour le web sémantique, mais il lui manque justement cette couche sémantique.

Listing 1 – Exemple de document XML

```
<?xml version="1.0"?>
<annuaire>
  <personne class="membre">
    <nom>Pelletier</nom>
    <prenom>Francois</prenom>
    <telephone>418-440-6904</telephone>
    <email>francois@francoispelletier.org</email>
  </personne>
</annuaire>
```

- ▶ RDF : Resource Description Framework
- ▶ Les relations entre des ressources identifiées par des URI sont représentées par des triplets (sujet, prédicat, objet) de la forme suivante :
<Ressource, Propriété, Valeur>.
- ▶ Le RDF ne contient pas de définition du vocabulaire. Il faut donc définir tous les termes utilisés dans un schéma RDFS.
- ▶ Le RDF Schema permet de définir le vocabulaire en utilisant les propriétés suivantes :
Class, Property, Type, subclassOf, Domain, Range. On voit que c'est très inspiré de la programmation orientée objet.

Listing 2 – Exemple de document RDF

```
<?xml version="1.0"?>
<rdf:RDF
xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
xmlns:pers="http://www.monannuaire.com/personne#">
<rdf:Description
  rdf:about="http://www.monannuaire.com/personne/FP">
  <pers:nom>Pelletier</pers:nom>
  <pers:prenom>Francois</pers:prenom>
  <pers:telephone>418-440-6904</pers:telephone>
  <pers:email>
    francois@francoispelletier.org
  </pers:email>
</rdf:Description>
</rdf:RDF>
```

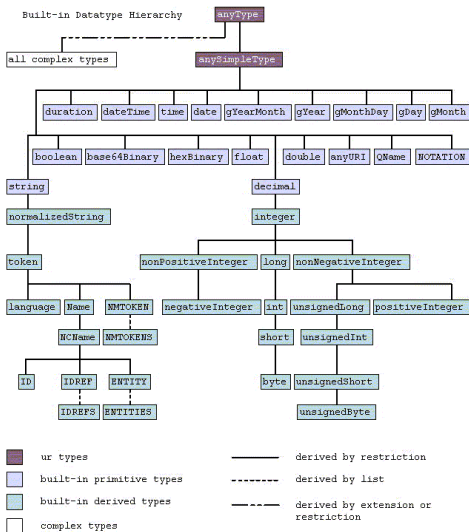


Figure – Type des données XML

OWL : Définir les ontologies

- ▶ Ontologie : Description de la taxonomie (vocabulaire) et de la structure des connaissances dans un domaine.
- ▶ OWL : Web Ontology Language
- ▶ Basé sur la logique descriptive.
- ▶ Définit des classes et des propriétés à utiliser dans un document RDF. C'est une généralisation de RDFS.

Listing 3 – Exemple de document OWL

```
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#"
  xml:base="http://www.monannuaire.com/">
  <owl:Class rdf:ID="Personne"/>
  <owl:DatatypeProperty
    rdf:ID="email">
    <rdfs:label
      rdf:datatype="http://www.w3.org/2001/XMLSchema#string">
      Adresse courriel de la personne
    </rdfs:label>
    <rdfs:range
      rdf:resource="http://www.w3.org/2001/XMLSchema#string"/>
    <rdfs:domain
      rdf:resource="#Personne"/>
    </owl:DatatypeProperty>
</rdf:RDF>
```

SPARQL

- ▶ SPARQL : SPARQL Protocol and RDF Query Language
- ▶ Permet de faire des requêtes de type clé-valeurs sur des données qui respectent la spécification RDF
- ▶ Fournit les opérateurs de fusion, de tri et d'agrégation : JOIN, SORT, AGGREGATE
- ▶ Fournit aussi des opérateurs de traversée de graphes
- ▶ Extension : GeoSPARQL

Listing 4 – Exemple de requête SPARQL

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?name
        ?email
WHERE
{
  ?person a foaf:Person .
  ?person foaf:name ?name .
  ?person foaf:mbox ?email .
}
```

Listing 5 – Exemple de requête SPARQL appelant un service

```
select * where {  
  service <http://fr.dbpedia.org/sparql> {  
    <http://fr.dbpedia.org/resource/Linux> ?p ?y  
  }  
}
```

Quelques projets du web sémantique

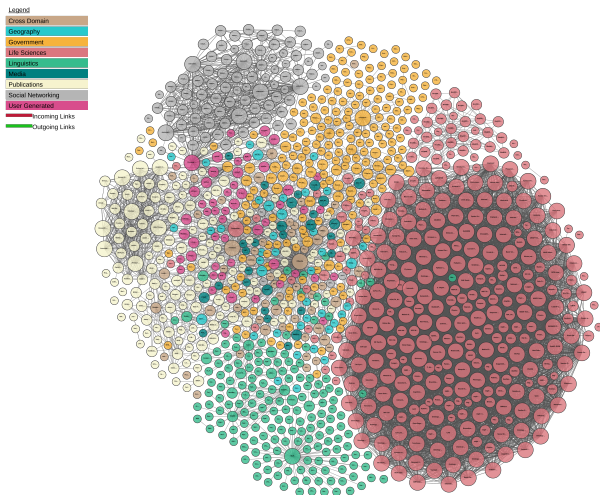


Figure – Linked Open Data Cloud

Quelques ressources sur les graphes

Bases de connaissances :

- ▶ DBPedia
- ▶ BabelNet

Ontologies et schémas

- ▶ Dublin Core
- ▶ schema.org

Répertoires

- ▶ Semantic Web Case Studies and Use Cases
- ▶ Datahub

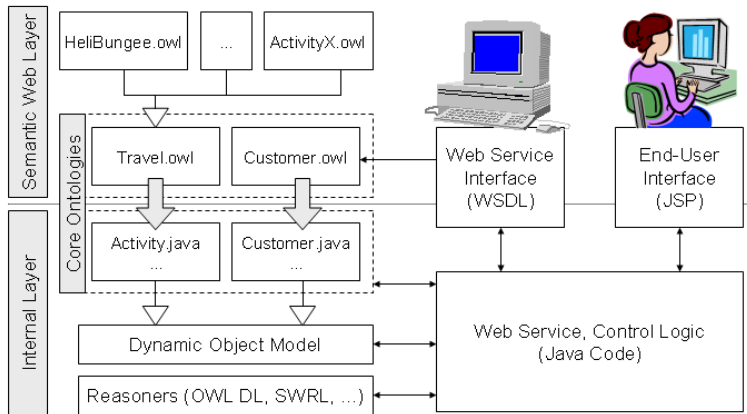


Figure – Construire des applications Source)

Logiciels libres pour explorer et développer le web sémantique

- ▶ Outils d'exploration :
 - ▶ CoReSe
 - ▶ Facete Logiciel utilisé par European Union Open Data Portal
 - ▶ Gephi
 - ▶ Structured Data Sniffer (Module pour navigateur)
- ▶ Outils de développement :
 - ▶ Open Semantic Search
 - ▶ Apache Jena
 - ▶ Stanford Protégé
 - ▶ Fedora Commons

Logiciels libres pour explorer et développer le web sémantique

- ▶ Bases de données (triple store) :
 - ▶ OpenLink Virtuoso
 - ▶ Jena TDB
 - ▶ Sesame (RDF4J)
- ▶ Bibliothèques et langages :
 - ▶ SWI Prolog
 - ▶ RDF 4 Haskell

Exemples pratiques

- ▶ Exemple R: Linked Open Piracy
- ▶ Exemple R: Données spatiales liées
- ▶ Music Genre - Force directed graph