

Introduction au WEB Sémantique

Cours 1 : Qu'est ce que le WEB sémantique ?

Odile PAPINI

POLYTECH

Université d'Aix-Marseille

odile.papini@univ-amu.fr

<http://odile.papini.perso.luminy.univ-amu.fr/sources/WEBSEM.html>

Plan du cours

- 1 Le Web aujourd'hui
- 2 Le Web Sémantique
- 3 Quelques repères historiques
- 4 Technologies du Web sémantique

Bibliographie I



Grigoris Antoniou & Frank van Harmelen

MIT university Press

Semantic web primer

<http://www.ics.forth.gr/isl/swprimer/presentation.htm>



Fabien Gandon, Catherine Faron-Zucker, Olivier Corby

INRIA Sophia Antipolis

le Web sémantique. DUNOD. 2012



David Genest

Université d'Angers

<http://www.info.univ-angers.fr/pub/genest/enseignement/index.html>



Michel Gagnon

Ecole Polytechnique Montréal

<http://www.professeurs.polymtl.ca/michel.gagnon/Publications/tutorielSWIG04.pdf>

Bibliographie II



Jérôme Euzenat
INRIA Grenoble

<https://hal.inria.fr/hal-00906620/document>



Le MOOC : Web sémantique et Web de données
INRIA Sophia-Antipolis
Fabien Gandon, Olivier Corby, Catherine Faron Zucker
<https://www.fun-mooc.fr/courses/course-v1:inria+41002+session03/about#>



W3C
World Wide Web Consortium : standards du Web
<http://www.w3.org/standards/semanticweb/>

Le Web aujourd'hui

Web

Succès d' Internet

- Documents
- Images
- Vidéos
- Applications
- Services
- Vie sociale
- Buisness
- ...



Figure: source : L. Temal. SFEIR

Le Web aujourd'hui

Web

Le Web aujourd'hui

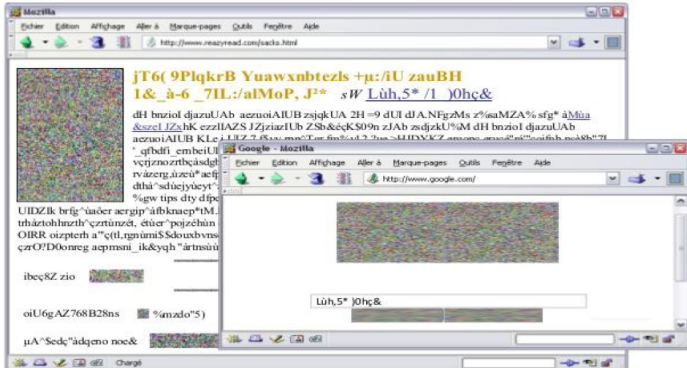
- information sur le Web essentiellement prévue pour être affichée (écran, imprimante) et lue par des humains
- il est essentiellement syntaxique : contenu quasi inaccessible aux traitements machines
- seuls les humains peuvent interpréter ces contenus

◀ ◻ ▶ ◀ ≡ ▶ ◀ ≡ ▶ ≡ ◻ ↺ 🔍 ↻

Le Web aujourd'hui

Web

Et pour la machine



Le Web aujourd'hui

Web

Le Web aujourd'hui

La machine stocke des informations
qu'elle ne peut pas interpréter



Le Web aujourd'hui

Web

Moteurs de recherche par mot clé

- Les activités Web ne sont pas particulièrement adaptés aux outils logiciels
 - A l'exception des moteurs de recherche par mot clé :
 - Google, altavista Yahoo ...
- Le Web n'aurait pas eu autant de succès sans les moteurs de recherche

Le Web aujourd'hui

Web

Problèmes des moteurs de recherche par mot clé

- faible précision
- résultats très sensibles au vocabulaire
- résultats : seulement pages Web
- intervention humaine pour interpréter et combiner les résultats
- résultats des recherches pas lisibles par d'autres logiciels

Le Web aujourd'hui

Web

Les principaux problèmes du Web d'aujourd'hui

- le sens des contenus Web n'est pas accessible aux machines :
manque de sémantique
- Exemple : difficulté de distinguer le sens du verbe **jouer** :
 - **elle joue des cymbales**
 - **elle joue des coudes**
 - **elle joue de son influence**

Le Web sémantique

Web

Le Web sémantique : objectif

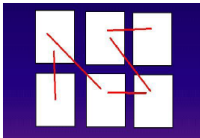
“The semantic Web is an extension of the current web in which information is given well-defined meaning, better enabling computers and people to work in cooperation”

Tim Berners-Lee, James Hendler, Ora Lassila
The Semantic Web, Scientific american, May 2001
<http://www.sciencificamerican.com>

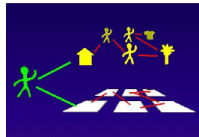
Le Web sémantique

Web

Le Web sémantique : objectif



Passer du web des liens
entre des pages peu
compréhensibles
par les machines à ...



un web de choses
reliées à la réalité et
compréhensibles par
les machines



Figure: source : T. Berners-Lee

Le Web sémantique

Web

Le Web sémantique : objectif

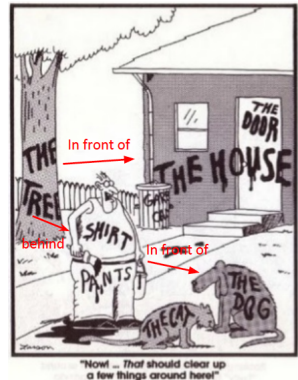
- le Web actuel :
 - pas de structure explicite globale
 - liens non exploitables sémantiquement
 - travail limité sur les informations
- le Web sémantique :
 - séparation de la forme et du contenu des informations
 - connaissances formalisées
 - lien sémantique entre informations
 - annotations plus riches
 - standard à base d'XML, mais ouverture

Le Web sémantique

Web

Le Web sémantique : objectif

- Identifier
- Expliciter
- Représenter



Le Web sémantique

Web

Le Web sémantique : Modélisation

Modéliser, représenter les connaissances :

Ontologies

- connaissances partagées que nous pouvons comprendre
- que la machine peut interpréter

Le Web sémantique

Web

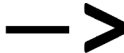
Le Web sémantique : Modélisation

Définition de l'ontologie

Ontologie



Philosophie: Branche fondamentale de la Meta - physique, qui s'intéresse à la notion d'existence, aux catégories fondamentales de l'existant et qui étudie les propriétés les plus générales de l'être .



ontologie



Informatique: Spécification explicite et formelle d'une conceptualisation d'un domaine de connaissance. [Gruber, 1993] .

Conceptualisation : Elaboration d'un *modèle abstrait* d'un domaine du monde réel en identifiant et en classant les *concepts pertinents* décrivant ce domaine .

Formalisation: Rendre cette conceptualisation exploitable par des machines .

11

Figure: source : L. Temal. SFEIR

Le Web sémantique : lier sémantiquement les informations

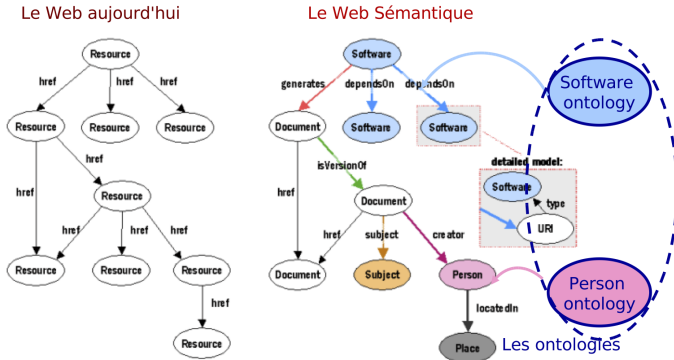


Figure: source : W3C Semantic Web Activity, Koivunen and Miller, 2001

Le Web sémantique

Ressources

Le Web est composé de ressources

- Chaque ressource a un URI
- URI (Universal Resource Identifier) : identifiant unique sur la toile (World Wide Web)
 - un URL (Universal Resource Locator)
 - ou tout autre identifiant unique

Le Web sémantique

Web

Le Web sémantique : impact sur gestion des connaissances

- gestion des connaissances : acquisition, accès, maintien des connaissances dans une organisation
- activité importante dans l'industrie
- importance pour des organisation internationales dispersées géographiquement
- la plupart des informations disponibles sont faiblement structurées (textes, sons, images, ...)

Le Web sémantique

Web

Limitation des technologies actuelles de gestion des connaissances

- **recherche d'information**
 - moteurs de recherche à base de mot-clé
- **extraction d'information**
 - intervention humaine nécessaire pour naviguer, chercher, interpréter, combiner
- **maintenance de l'information**
 - incohérences de terminologie, information dépassée
- **visualisation de l'information**
 - impossible de définir des vues sur la connaissance Web

Le Web sémantique

Web

Le Web sémantique adapté à la gestion des connaissances

- les connaissances sont organisées en espaces conceptuels selon leur signification
- outils automatiques pour la maintenance et la découverte de connaissances
- réponse à des questions sémantiques
- réponse à des questions sur plusieurs documents
- possibilité de définir qui peut voir certaines parties de l'information

Le Web sémantique

Quelques applications du Web sémantique

Réseaux sociaux, blogs, plateformes communautaires

- agréger plusieurs données entre elles



...

Le Web sémantique

Quelques applications du Web sémantique

Réseaux sociaux, blogs, plateformes communautaires

The screenshot shows the CC Search website. At the top is the Creative Commons logo. Below it is a green banner with the text "Try the new CC Search beta, with list-making and one-click attribution!". The main search area features a green "CC Search" button with the text "Find content you can share, use and remix". To the right is a search input field containing "eiffel tower". Below the input field are two checkboxes: "I want something that I can..." with options "use for commercial purposes;" and "modify, adapt, or build upon." Below the search area is a section titled "Search using:" with a grid of service icons: Europeansia Media, Flickr Image, Google Web, Google Images, Jamendo Music, Open Clip Art Library Image, SpinXpress Media, Wikimedia Commons Media, YouTube Video (highlighted in blue), Pixabay Image, ccMixer Music, and SoundCloud Music. At the bottom left is a disclaimer: "Please note that search.creativecommons.org is not a search engine, but rather offers convenient access to search services provided by other independent organizations. CC has no control over the results that are returned. Do not assume that the results displayed in this search portal are under a CC license. You should always verify that the work is actually under a CC license by following the link. Since there is no registration to use a CC license, CC has no way to determine...". At the bottom right is a yellow box with the text "Add CC Search to your browser." and a link "Learn how to switch to or from CC Search in your Firefox search bar." Below this is a language selector showing "English" and a "Help Translate" link. At the very bottom right are navigation icons.

creative commons

Try the new CC Search beta, with list-making and one-click attribution!

CC Search
Find content you can share, use and remix

eiffel tower

I want something that I can... ☒ use for commercial purposes;
☒ modify, adapt, or build upon.

Search using:

Europeana Media	Flickr Image		Google Web
Google Images Image	Jamendo Music	Open Clip Art Library Image	SpinXpress Media
Wikimedia Commons Media	YouTube Video	Pixabay Image	ccMixer Music
SoundCloud Music			

Please note that search.creativecommons.org is not a search engine, but rather offers convenient access to search services provided by other independent organizations. CC has no control over the results that are returned. Do not assume that the results displayed in this search portal are under a CC license. You should always verify that the work is actually under a CC license by following the link. Since there is no registration to use a CC license, CC has no way to determine...

Add CC Search to your browser.
Learn how to switch to or from CC Search in your Firefox search bar.

English Help Translate

Le Web sémantique

Quelques applications du Web sémantique

Recherche, classement bibliographique, documentaire

- recherches simultanées

The screenshot shows the BnF Catalogue général search interface. At the top, there's a header with the BnF logo and the text 'Catalogue général'. Below this is a search bar with a dropdown menu set to 'Tout' and a search button. To the right of the search bar are links for 'Recherche avancée', 'RECHERCHES CIBLÉES', and 'DANS L'UNIVERS'. Below the search bar, there's a section for 'Recherche avancée' with a description: 'La recherche avancée permet une interrogation plus fine du catalogue général de la BnF. Tous les critères présents sur cette page peuvent être utilisés de manière indépendante ou croisée. Plus d'informations dans la rubrique Aide ?'. Below this is a section for 'Par typologie' with a list of document types: Document cartographique, Document iconographique, Audiovisuel, Musique, Objets monétiformes, and Périodiques. At the bottom right, there are navigation icons for back, forward, and search.

{BnF Catalogue général

Espace personnel Aide Une question ? Historique

Tout Recherche avancée RECHERCHES CIBLÉES DANS L'UNIVERS

Accueil / Recherche avancée

Recherche avancée

La recherche avancée permet une interrogation plus fine du catalogue général de la BnF. Tous les critères présents sur cette page peuvent être utilisés de manière indépendante ou croisée. Plus d'informations dans la rubrique Aide ?

Effacer la recherche Laisser la recherche

Par mots

Par localisation

Par nature de document

Par typologie

- Document cartographique
- Document iconographique
- Audiovisuel
- Musique
- Objets monétiformes
- Périodiques

Le Web sémantique

Web

Le Web sémantique : impact sur le commerce électronique (B2C)

- un scénario typique : les utilisateurs visitent des sites de magasins en ligne, étudient leurs offres sélectionnent et commandent des produits
- activité importante dans l'industrie
- idéalement, les utilisateurs devraient visiter tous, ou une grande partie des sites des magasins MAIS cela prend trop de temps !
- les “shopbots” (robots logiciels) : outils utiles

Le Web sémantique

Web

Limitation des “shopbots”

- reposent sur des “wrappers”(conteneurs à contenu caché) nécessitant une programmation importante
- les “wrappers” doivent être reprogrammés lorsqu’un magasin change ses outils
- les “wrappers” extraient de l’information sur la base d’une analyse textuelle
 - propension aux erreurs
 - information extraite limitée

Le Web sémantique

Web

Le Web sémantique adapté commerce électronique (B2C)

- agents logiciels l'information sur le produit et les termes du service
 - prix, information sur le produit, livraison, politique de confidentialité seront interprétés et comparés aux besoins de l'utilisateur
- informations sur la réputation des magasins
- agents acheteurs sophistiqués pourront conduire des négociations

Le Web sémantique

Web

Le Web sémantique : impact sur le commerce électronique (B2B)

- la plus grande promesse économique
- actuellement repose la plupart du temps sur des EDI
(Electronic Data Interchange)
 - technologie seulement comprise par des experts
 - difficulté de programmation, de maintenance, error-prone
 - programmation séparée pour chaque communication B2B
- le Web semble être une parfaite infrastructure
 - mais B2B mal géré par les standards web

Le Web sémantique

Web

Le Web sémantique adapté commerce électronique (B2B)

- enregistrement des partenariats sans charges indirectes
- différences entre terminologies résolues par l'utilisation de modèles de domaine abstrait standards
- échange de données par l'utilisation de service de translation
- enchères, négociations, ébauche de contrats automatiquement (ou semi-automatiquement) réalisés par des agents logiciels

Le Web sémantique

Quelques applications du Web sémantique

Commerce, e-commerce

- description de manière structurée les produits, les prix les informations sur l'entreprise . . .

Best Buy Magasinez English Magasins

Localisateur de magasins Best Buy

Recherche par code postal ou par ville et province

Montréal, QC, Canada

Filtrer par service (0)

10 magasins près de "Montréal, QC"

- 1 Best Buy Mobile Montreal Eaton Centre** 0.28 km
Open Today Until 21:00
705 Rue Sainte-Catherine O, Unit 2114A
Montréal, QC H3B4G5
(514) 798-8997
[Get Directions](#) [Détails du magasin](#)
- 2 Best Buy Centreville** 0.34 km
Open Today Until 21:00
470 Rue Ste. Catherine O
Montréal, QC H3B1A6
(514) 919-2600

Le Web sémantique

Quelques applications du Web sémantique

Web de données (Linked Open Data)

- Utiliser les URIs pour nommer les ressources
- Utiliser des URIs HTTP pour pouvoir obtenir des représentations de ces ressources
- Fournir ces représentations en utilisant les langages et protocoles standards (RDF, SPARQL, ...)
- Inclure des liens pour permettre de découvrir de nouvelles données

Le Web sémantique

Quelques applications du Web sémantique

Web de données (Linked Open Data)

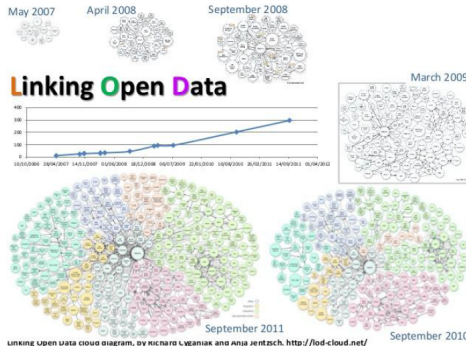


Figure: source : L. Temal SFEIR

Le Web sémantique

Quelques applications du Web sémantique

Web de données (Linked Open Data)

Linked Open Data

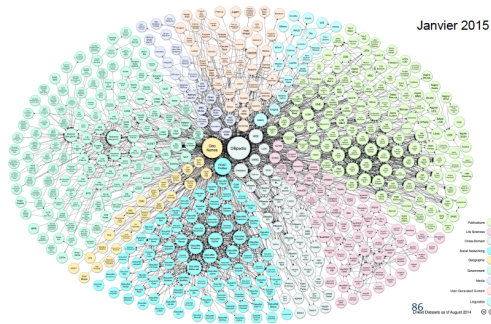


Figure: source : L. Temal SFEIR

Le Web sémantique

Quelques applications du Web sémantique

Web de données (Linked Open Data)

juin 2018

The [Linked Open Data](#) Cloud

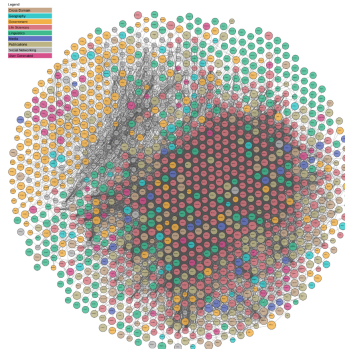


Figure: source : <https://lod-cloud.net/>

Quelques repères historiques : 1945 MEMEX

Collective Memory

- "As We May Think" published in July 1945 by Vannevar Bush.
- Transform an information explosion into a knowledge explosion.
- "memex" a machine that achieves a higher level of organized knowledge (like human memory processes).

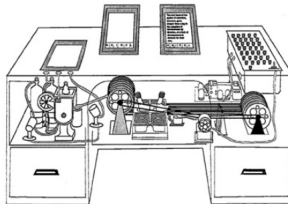


Figure: source : Z. Bouraoui. Université d'Artois

Quelques repères historiques : 1965 Hypertexte

HyperText

- Theodor Holm Nelson invented the terms "hypertext" and "hypermedia" (published in 1965).
- "Let me introduce the word 'hypertext' to mean a body of written or pictorial material interconnected in such a complex way that it could not conveniently be presented or represented on paper".



Hypertext Editing System—Brown University, 1969

Figure: source : Z. Bouraoui. Université d'Artois

Quelques repères historiques : 1990 World Wide Web (1)

WWW : World Wide Web

- Proposed by Tim Berners-Lee, a scientist at CERN.
- Information-sharing facility, among physicists working at CERN and other academic institutions.
- Berners-Lee is the director of the World Wide Web Consortium (W3C).

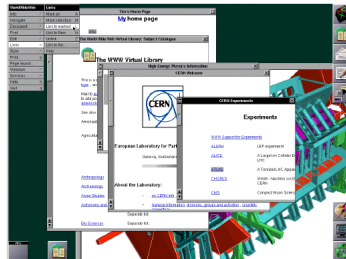
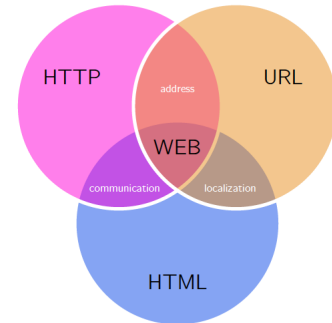


Figure: source : Z. Bouraoui. Université d'Artois

Quelques repères historiques : 1990 World Wide Web (2)

WWW : 1990



HTML Hypertext Markup Language

URL Uniform Resource Locator

HTTP Hypertext Transfer Protocol



Figure: source : Z. Bouraoui. Université d'Artois

Quelques repères historiques : 1994 W3C

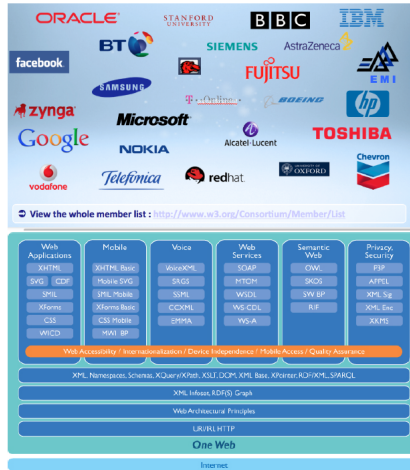
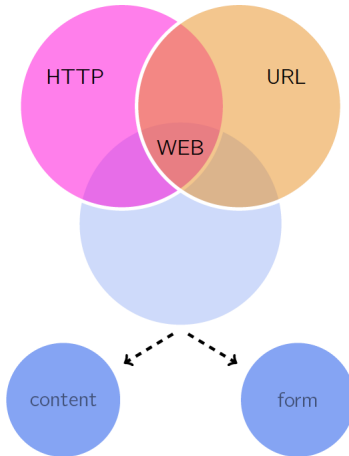


Figure: source : Z. Bouraoui. Université d'Artois

Quelques repères historiques : 1996 CCS

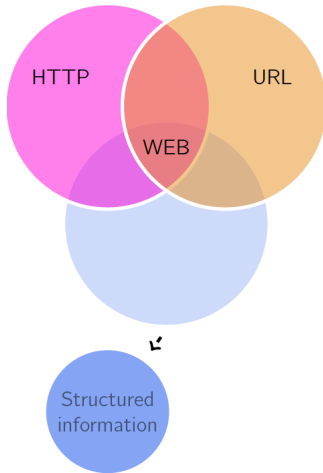


CSS Cascading Style Sheets



Figure: source : Z. Bouraoui. Université d'Artois

Quelques repères historiques : 1998 XML



XML Extensible Markup Language

XSLt Transforming into other one, or other formats.

XPath Selecting nodes

XPointer Addressing components.

XLink Creating internal and external links

XQuery Querying

XProc Pipelining treatments.

Figure: source : Z. Bouraoui. Université d'Artois

Le Web sémantique

Web

Technologies du Web sémantique

- Méta données explicites
- Ontologies
- Logique et raisonnement
- agents

Le Web sémantique

Web

Méta données explicites

- représentation plus facilement exploitable par les machines
- méta données : données sur les données
- méta données capturent une partie de la signification des données
- le Web sémantique ne repose pas sur des manipulations basées sur du texte mais plutôt sur des méta données exploitables par des machines

Le Web sémantique

EXEMPLE HTML (G. Antoniou, F. van Harmelen)

```
<h1>Centre de kinésithérapie Agilitas </h1>
Bienvenue à la page d'accueil du Centre de kinésithérapie
Agilitas. Ressentez-vous de la douleur? Avez-vous eu un
accident? Notre personnel
Lise Davanport,
Josiane Bouville (notre charmante secrétaire) et
Etienne Matthieu vont prendre soin de vous.
<h2>Horaire des consultations</h2>
Lun 11.00 - 19.00<br>
Mar 11.00 - 19.00<br>
Mer 15.00 - 19.00<br>
Jeu 11.00 - 19.00<br>
Ven 11.00 - 15.00<p>
Veuillez noter que nous n'avons pas de consultations les
semaines de
<a href=". . .">State Of Origin</a> games.
```

Le Web sémantique

EXEMPLE XML (G. Antoniou , F. van Harmelen)

<société>

<traitementProposé>Kinésithérapie</traitementProposé>

**<nomSociété>Centre de Kinésithérapie
Agilitas</nomSociété>**

<personnel>

<kiné>Lise Davanport</kiné>

<kiné>Etienne Matthieu</kiné>

<secrétaire>Josiane Bouville</secrétaire>

</personnel>

</société>

Le Web sémantique

Web

Ontologies

- en philosophie :
 - **étude de la nature de l'existence**
- en informatique :
 - **spécification formelle et explicite d'une conceptualisation**
 - structurées en termes de concepts et de relations entre concepts
 - formalisation partagée sur un domaine
 - utiles pour l'organisation et la navigation sur les sites Web

Le Web sémantique

Web

Exemple d'ontologie

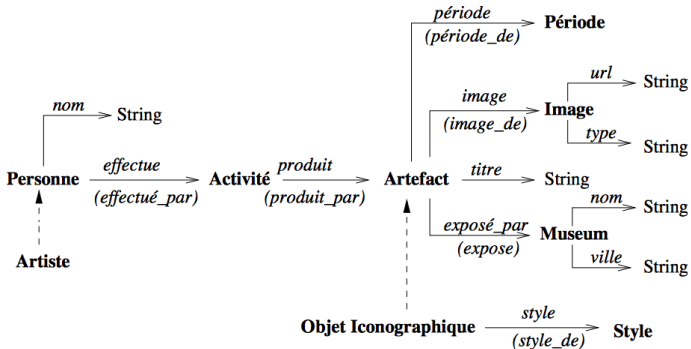


Figure: source : B. Amann INRIA

Le Web sémantique

Web

Langages de représentation des ontologies

- RDF (Resource Description Framework)
- RDFS (RDF Schéma)
- OWL (Ontology Web Language)

Le Web sémantique

RDF

- Resource : Documents, livres, entreprises, services, media, ...
- Description : Relations, attributs, relations entre les ressources
- Framework : Langages et syntaxes pour ces descriptions

Atome de connaissance en RDF : triplet (Sujet, Predicat, Objet)

Représentation sous forme de graphe : graphe RDF

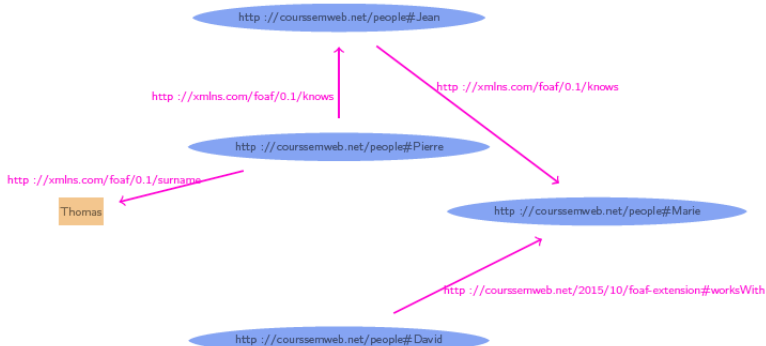


Le Web sémantique

RDF : exemple

Pierre knows Jean
Jean knows Marie

Pierre's surname is Thomas
David works with Marie



Le Web sémantique

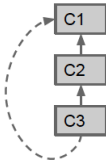
RDFS

RDF Schema : langage de description des ontologies légères

- recommandation du W3C depuis 2004
- permet de nommer :
 - les classes
 - les propriétés
- permet de définir une organisation hiérarchique des classes et propriétés
- méta-modèle : donne une sémantique au graphe RDF

Le Web sémantique

RDFS exemple : transitivité dans la hiérarchie des classes inférence de nouveaux triplets



si (c2, subClassOf, c1)
et si (c3, subClassOf, c2)
alors
(c3, subClassOf, c1)

Le Web sémantique

OWL

OWL : Ontology Web Language
langage de description des ontologies lourdes

- OWL 1 recommandation du W3C depuis 2004
- OWL 2 recommandation du W3C depuis 2009
- plus expressif que RDFS permet d'exprimer
 - équivalence de classes/propriétés
 - égalité de ressources, différence, de contraire, de symétrie, cardinalité, ...

Le Web sémantique

Web

Interrogation

SPARQL (Sparql Protocol And RDF Query Language)

- langage d'interrogation de graphe RDF
- format des résultats des requêtes
- protocole d'accès aux données RDF

Le Web sémantique

Web

Logique et raisonnement

- logique : étude des principes et formes du raisonnement
- langages formels de représentation des connaissances (ex Logiques de description)
- sémantique formelle bien définie
- déduction automatique : outils de raisonnement (ex PELLET)
- compromis entre expressivité et complexité calculatoire

Le Web sémantique

Web

Technologies du Web sémantique

- Méta données
 - identification et extraction de l'information des sources Web
- Ontologies
 - recherches sur le web, récupération de l'information interprétée
 - communication entre agents
- Logique
 - traitement de l'information récupérée, inférence de conclusions

Le Web sémantique : Approche par couches

Web

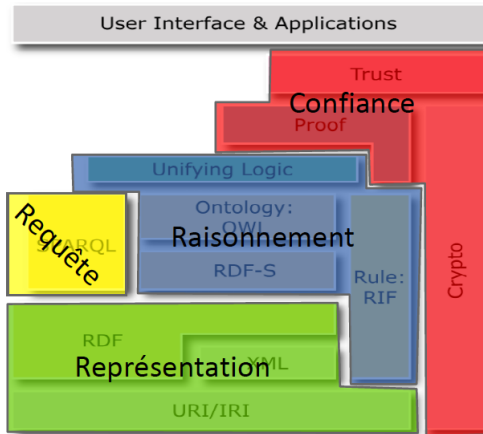
Le développement du web sémantique s'opère par étapes

- chaque étape construit une couche au dessus d'une autre

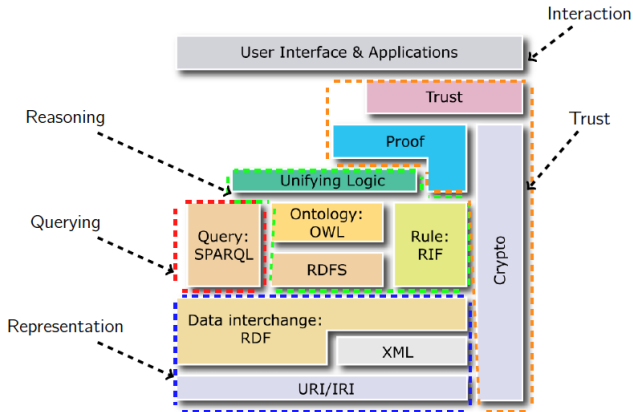
principes

- compatibilité descendante
- compréhension partielle ascendante

Le Web sémantique : Approche par couches



Le Web sémantique : Approche par couches



Le Web sémantique : Approche par couches

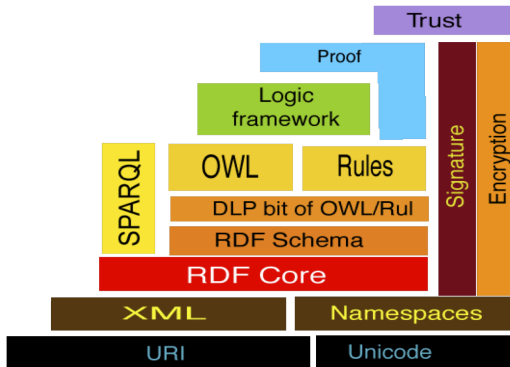
- couche XML
 - base syntaxique
- couche RDF
 - RDF : modèle de données basique pour les faits
 - RDF Schéma : langage pour les ontologies
- couche Ontologie
 - langage plus expressif que RDF Schéma
 - standard courant pour le web : OWL

Le Web sémantique : Approche par couches

- **couche Logique**
 - évolution des langages pour les ontologies
 - applications spécifique pour des connaissances déclaratives
- **couche Contrôle**
 - génération de contrôles, validation
- **couche Sécurisation**
 - signatures numériques
 - recommandations,...

Le Web sémantique : Approche par couches alternative

Le web sémantique : structuration



Le Web sémantique : Approche par couches alternative

- prend en compte les développements récents :
- les différences essentielles :
 - couche ontologie : 2 langages standard pour les ontologies pour le web : OWL et un langage basé sur des règles
 - DLP : intersection de OWL et la logique basé sur les clauses de Horn
- l'architecture du web sémantique est en débat : elle va subir des modifications et évoluer dans le futur