

Le web sémantique (2ème partie) : RDF, RDFs, OWL, etc. De la panoplie du Web sémantique...

[Retour au sommaire de la lettre](#)

Le Web sémantique a longtemps été perçu comme une vue de l'esprit, inadapté aux contingences actuelles du Web, incompréhensible par l'internaute lambda. A présent, tous les grands acteurs du Web ont mis à profit sa technologie dans des applications intelligentes et utiles. Quelle panoplie est mise en œuvre pour le déploiement du Web sémantique ou Web of data ? Loin d'un simple panorama des outils de représentation des connaissances, l'intérêt pour nous est de comprendre comment ils fonctionnent et à quoi ils servent. Il s'agit d'outils conçus selon une architecture proche du mode de cognition humaine, basés à la fois sur l'intelligence artificielle et la linguistique. Exploration...

Nous avons vu dans la première partie de cette série d'articles (voir lettre R&R de septembre 2008) les ressorts du Web sémantique, sa vocation, ses initiateurs. Cette deuxième partie est destinée à présenter les outils qui participent au déploiement du Web sémantique.

Définitions du web sémantique :

Le Web sémantique défini **en théorie** : "un web compréhensible par des machines".

Le web sémantique défini **en pratique** : "une infrastructure composée de langages pour exprimer la connaissance, pour décrire les relations entre connaissances, pour décrire les conditions d'utilisation, pour décrire les garanties et les modes de paiement et de dispositifs permettant de trouver les ressources."

De ces deux définitions, la seconde est la plus parlante, car c'est en termes de *représentation des connaissances* qu'il faut penser le Web sémantique et le Web de demain. Longtemps, on a reproché au Web sémantique d'être trop "ésotérique" et ses applications éventuelles ne relevaient pas de l'évidence.

A présent, les applications existent et, surtout, la prétention du "Web de données" est de relever des défis dans lesquels l'internaute *lambda* puisse se retrouver.

Quels sont les besoins ?

- Une *identification* et un *accès* aux ressources du Web de façon simple, robuste, et efficace (URI).
- Des langages pour décrire le *contenu* des documents/ressources et des moteurs d'inférences pour les manipuler...
- Des *ressources* de plusieurs types : ontologies (bases de connaissances), bases de données, éléments de code, etc.

Pour quoi faire ?

- Créer des liens automatiquement entre documents en ligne (de même thématique ou de thématique "sœur" ou "cousine")
- Créer des liens entre personnes et organisations (FOAF : *Friend Of A Friend*).
 - Ex : un document qui parle de "Barack Obama" va être directement lié à un document qui parle de "politique américaine" sans jamais citer Barack Obama. Par définition et de par son profil explicité au travers de métadonnées compréhensibles par la machine, Barack Obama est forcément associé à la politique américaine.
- Faire des références automatiques à des événements (iCalendar).
 - Ex : un document qui parlera de Barack Obama sera lié à un document qui parlera du 04 novembre 2008.
- Engendrer des fils (RSS)
- Retrouver des documents de manière simple et pertinente

Que faut-il faire et comment pour s'engager dans une démarche "Semantic Web"?

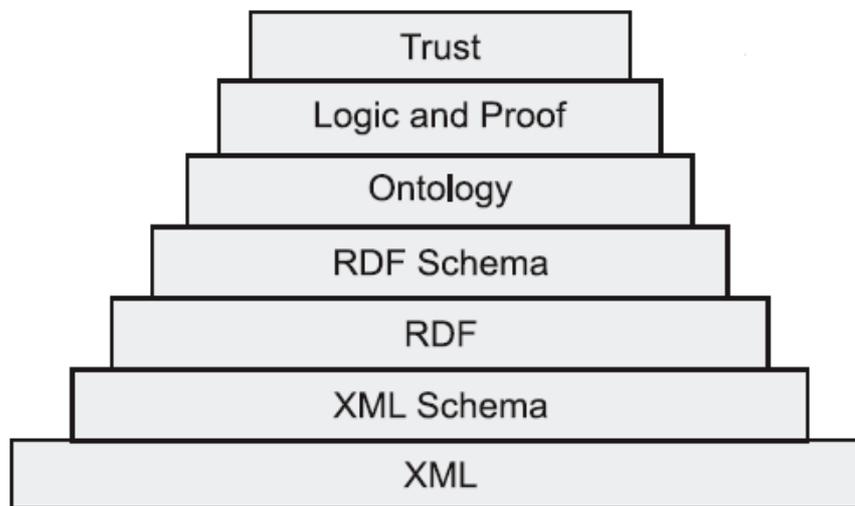
Décrire ses données en :

- Tirant parti des autres ressources (quelles qu'elles soient),
- Etendant ses ressources,
- Les rendant disponibles c'est-à-dire bien décrites pour quelles soient utilisables par les machines.

Le Web sémantique peut donc se résumer à ce simple fait : **la description des données**. C'est en décrivant ses **données** selon une architecture appropriée qu'elles seront plus accessibles et pourront être gérées de manière optimale. A présent, voyons donc quelle est cette "architecture appropriée".

Le "semantic Web layer cake" ou "le gâteau du Web sémantique"

Le W3C a été l'un des leaders dans le développement de technologies Web. Cet organisme a également été l'un des moteurs de la promotion du Web sémantique. L'approche de l'évolution du Web Sémantique est souvent décrite à travers la représentation du fameux "*Semantic Web cake*". Souvent reproduit, il illustre les différentes couches de l'architecture du Web sémantique.



Chaque couche est construite en fonction de la couche inférieure. Il permet de définir un petit dictionnaire des outils du Web sémantique :

- **XML** : *Extensible Markup Language* a été utilisé pour définir tout nouveau langage utilisé pour échanger des données sur le Web.
- **XML Schema** : utilisé pour décrire la structure de langages XML.
- **RDF** : *Ressource Description Framework*. Un langage flexible capable de décrire toute forme d'information et de métadonnées. "Dire quelque chose à propos de quelque chose".
- **RDF Schema** : Un cadre qui fournit un moyen de préciser les vocabulaires de base pour l'application spécifique RDF à utiliser.
- **Ontologie** : langage utilisé pour définir les mots et termes spécifiques à un vocabulaire. OWL est un langage d'ontologie conçu pour le Web sémantique.
- **Logic and Proof**: il s'agit d'un raisonneur basé sur une logique qui permet d'inférer des connaissances à partir de connaissances déjà établies et explicitées. L'exemple le plus fréquent est le syllogisme. Prenons le célèbre syllogisme issu de la logique aristotélicienne :

Tous les hommes sont mortels. Tous les grecs sont des hommes. Tous les grecs sont mortels.
Les deux premières propositions (ou prémisses) sont des connaissances connues et établies. La troisième proposition (ou conclusion) est inférée à partir des deux premières.

De la même manière, à partir de connaissances explicites, le raisonneur va pouvoir tirer des conclusions. Dans le cas concret du Web, on peut imaginer la connaissance suivante :

Barack Obama est président des Etats-Unis. Tout président des Etats-Unis habite à la Maison Blanche. Barack Obama habite à la Maison Blanche.

La dernière proposition est inférée à partir des deux premières.

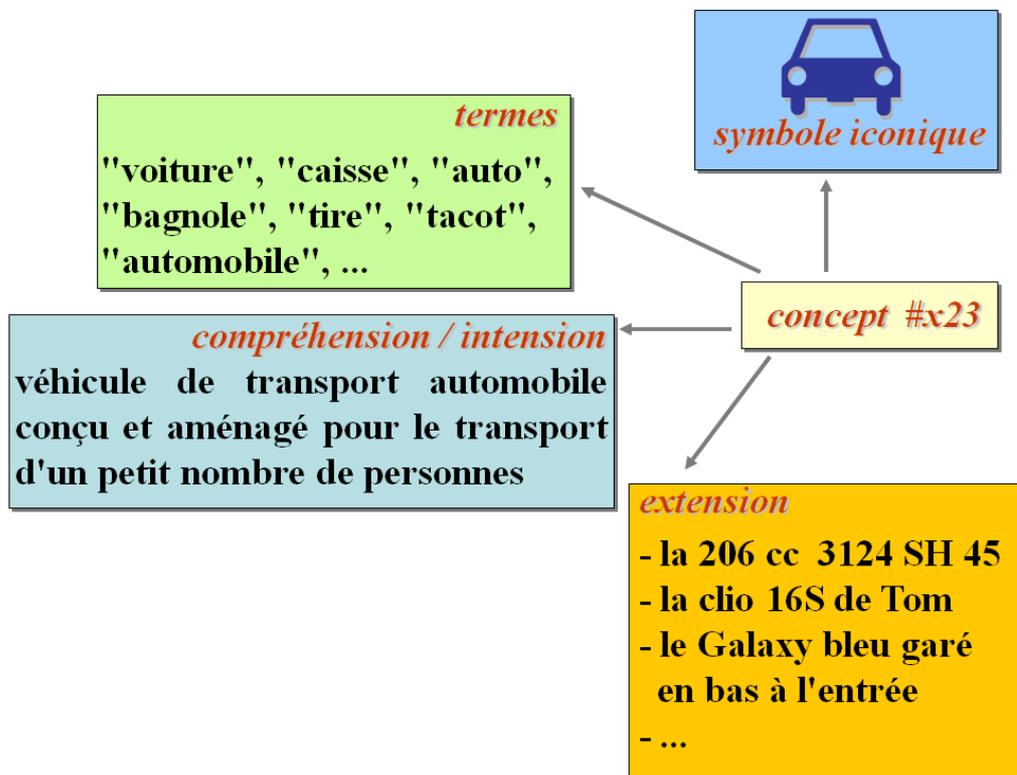
- **Trust:** Un moyen d'assurer l'authentification et la fiabilité des données, des services et des agents.

L'ordinateur et l'humain

Ces outils vont permettre de réaliser ce à quoi est voué le Web sémantique : "sémantiser" les informations, de sorte qu'elles soient comprises non seulement par les humains mais aussi par les machines.

Petite démonstration :

Représentation de ce que l'humain sait, comprend et interprète : le concept de "voiture"



Cette connaissance pour l'humain est informelle (non mise en forme). Pour qu'elle soit compréhensible par une machine, elle doit être mise en forme (formelle).

Autre démonstration :

Ce que l'on écrit sur le Web :

"Brioude-Internet-Référencement a écrit un article intitulé 'De la panoplie du web sémantique' "

Ce que l'ordinateur comprend :

Blablablablablablaba texte blablablablablaba

Ce que l'ordinateur devrait comprendre grâce au Web sémantique :

> *http:// www.abondance.com/recherche-referencement/2008/11/la-panoplie-du-web-semantique.html/ est un article*

> *Un article est un document*

> *http:// www.abondance.com/recherche-referencement/2008/11/la-panoplie-du-web-semantique.html/ parle de Web sémantique*

> *l'auteur de cet article est Brioude-Internet-Référencement (BIR)*

La simple phrase "Brioude-Internet-Référencement a écrit un article intitulé 'De la panoplie du web sémantique'" suggère pour un être humain un mode de compréhension implicite se référant à des ontologies – des descriptions des choses du monde - acquises au cours de la construction langagière et cognitive. A savoir, un article est un contenu textuel de l'ordre du document ; l'article en question parle de Web sémantique ; l'auteur de l'article est Brioude-Internet-Référencement qui est une entité (on pourrait rajouter une entreprise).

Constat : Toutes les connaissances que nous venons d'explicitier se formulent de la même manière :

X est Y (Un article est un document/ Brioude-Internet-référencement est une entité)

Ce mode de formulation (sujet/ verbe /complément) a été repris tel quel pour formaliser la connaissance dans les outils du Web sémantique : c'est la vocation du RDF.

RDF est un modèle de triplets c'est-à-dire que toute connaissance est décomposée en **sujet / prédicat / objet**. (sujet/verbe/complément)

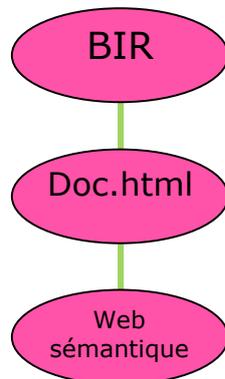
Exemples :

Doc.html / a pour auteur / BIR

Doc.html / a pour sujet / le Web sémantique

Les triplets peuvent aussi être vus comme les arcs d'un graphe RDF:

sommet / arrête / sommet



BIR est auteur de doc.html (qui) a pour sujet Web sémantique.

La connaissance est ici formalisée en un graphe, à savoir en **sujet/prédicat /objet (sommet/arrête/sommet)**:

- Les sommets sont les contenus textuels (BIR, doc.html, web sémantique).
- Les arêtes ou arcs sont la connaissance supplémentaire apportée à ce contenu textuel (est auteur, a pour sujet).

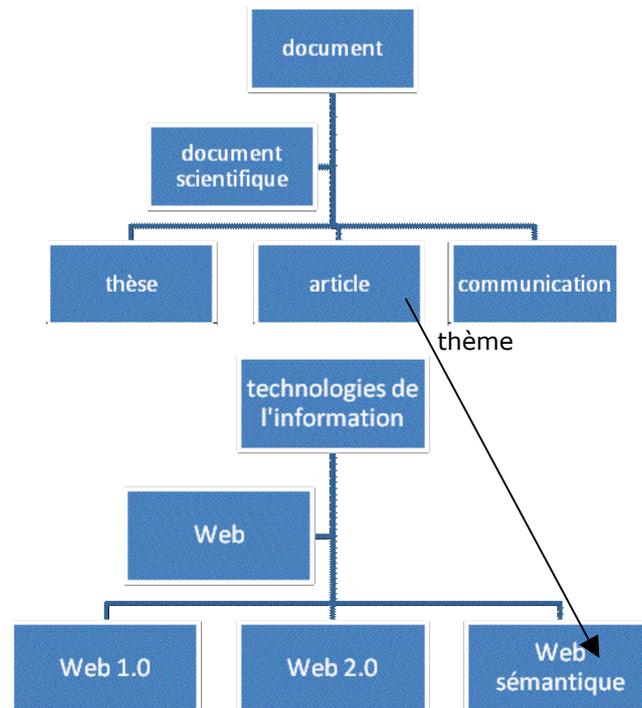
Mais à quoi bon apporter de la connaissance à l'ordinateur ?

Les connaissances ainsi formalisées, les graphes RDF pourront s'additionner en un vaste graphe mondial qui représentera la connaissance universelle.

Comment relier ces graphes ?

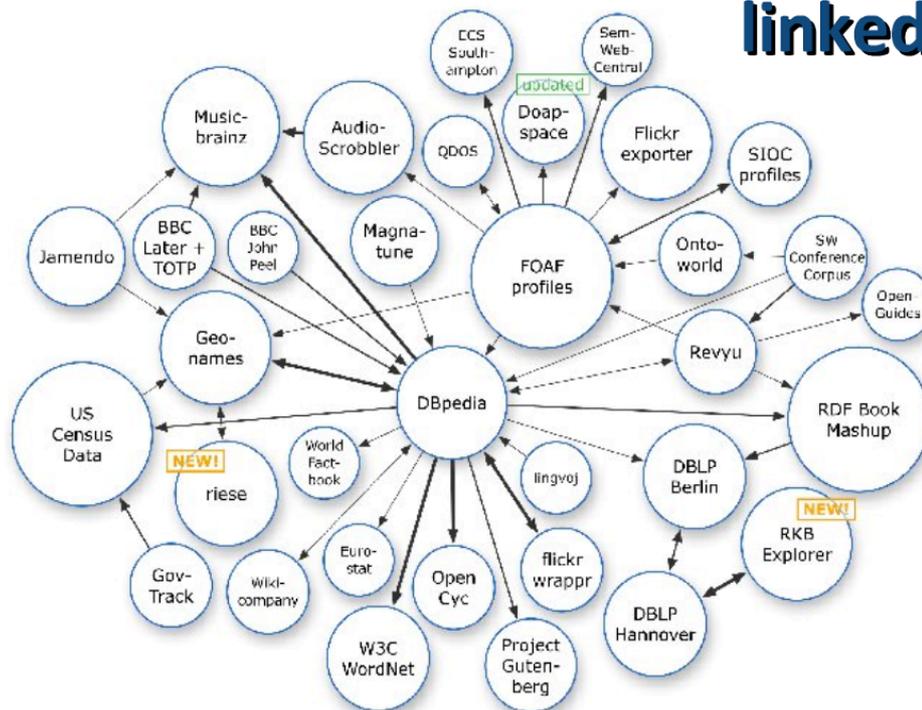
Avec une même description et un vocabulaire commun pour expliciter les connaissances, il sera possible de relier ces graphes. Si tel article parle de Web sémantique alors il pourra être relié automatiquement à un autre article parlant lui aussi de Web sémantique. Simplement en se basant sur les triplets RDF. Et pour aller plus loin, il pourra être relié à tout document parlant de Web ou de technologies de l'information. En effet les ontologies permettront de révéler les liens de "famille" entre choses.

La machine peut ainsi déduire qu'un article qui parle de "Web sémantique" est aussi un document qui parle de "Technologies de l'information".



Ainsi d'ontologies en ontologies, toutes les connaissances peuvent être reliées entre elles : des personnes (FOAF), des médias, des connaissances encyclopédiques (Wikipédia), des entités géographiques, des événements, ... pour arriver à une seule énorme ontologie.

linked data



Nous en arrivons donc au point crucial : **les effets sur les résultats de recherche et sur le mode de recherche.**

Pour une recherche du type "*document technologies de l'information*" ou "*article Web*", cet article sur le "Web sémantique" ressortira sur les pages de résultats SANS JAMAIS AVOIR MENTIONNÉ LES MOTS(-CLES) : document, Web, article, technologies de l'information ! Intéressant, non ?

En outre, les requêtes peuvent être beaucoup plus étoffées que la simple formulation de mots-clés. Quand bien même la recherche "plein texte" par mots-clés est une pratique rentrée dans les mœurs, d'autres méthodes sont possibles avec la formalisation des données dans le Web sémantique :

- Les requêtes structurées : "livres sur le Web sémantique".
- Les requêtes renseignées : "autobiographie de Tim Berners-Lee".
- Les requêtes composites : "quelle bibliothèque la plus proche peut me prêter un livre de Tim Berners-Lee sur le web sémantique ?".

Les documents ainsi renseignés pourront être d'autant **plus indexables** par les moteurs de recherche et seront d'autant **plus pertinents** pour l'internaute.

La suite de notre série sera consacrée aux applications du web sémantique : quels sont les acteurs du web qui utilisent ouvertement ou non les technologies du web sémantique et dans quels buts. Rendez-vous dans deux mois !

Myriam MALORON

veille@brioude-internet.fr

[Brioude Internet Référencement](http://brioude-internet.fr)

Réagissez à cet article sur le blog des abonnés d'Abondance :

<http://abonnes.abondance.com/blogpro/2008/11/le-web-smantique-2me-partie-rdf-rdfs.html>