## A la découverte du Web sémantique (1ère partie)

Retour au sommaire de la lettre

*Dans cette série de quatre articles, nous allons essayer d'explorer le monde du Web Sémantique, sur lequel travaille actuellement tous les grands acteurs de la recherche d'information. Tout d'abord, nous allons, dans cette première partie, nous attacher surtout à bien définir ce que recouvre le concept de Web Sémantique, à la croisée de diverses disciplines telles que l'intelligence artificielle, la linguistique, le web, les sciences de l'information, afin de bien comprendre la façon dont il va être intégré dans les outils de recherche d'un futur qui peut s'avérer bien plus proche qu'on ne le pense parfois...*

Annoncé comme le futur du Web il y a déjà une décennie par Tim Berners Lee, son père fondateur, le WS (pour Web Sémantique), aussi appelé "Web of data", commence à sortir des fins fonds des laboratoires universitaires pour être intégré dans diverses applications et chez de multiples acteurs du Web (Google et Yahoo pour ne citer qu'eux) sans roulements de tambours ; et pourtant, le Web sémantique est bien appelé à révolutionner le monde du *search* et du Web par extension. Pour comprendre les démarches et enjeux liés à cette notion, nous publierons une série d'articles tentant de faire le tour d'horizon du Web sémantique, un éclairage qui passera par une analyse à la fois historique, technique et conceptuelle et par la mise au jour des conséquences à venir sur la pratique du Web.

Notre série se composera des 4 articles suivants:
*1. Introduction au Web sémantique "I have a Dream..."*
*2. RDF, RDFa, RDFs, ontologies, OWL, etc ou De la panoplie du Web sémantique*
*3. Les applications du web sémantique : qui, quoi, pour quoi faire ?*
*4. Les conséquences sur la pratique du Web*
Ils seront publiés dans la lettre Recherche et Référencement dans les prochains mois.

### *Introduction au Web sémantique : "I have a Dream..."*

La spécificité du Web sémantique réside dans les multiples influences dont il fait l'objet. Il est en effet à la croisée de diverses disciplines telles que l'intelligence artificielle, la linguistique, le web, les sciences de l'information pour l'essentiel.

**Historique**

L'acte fondateur du web sémantique est la parution de l'article de Tim Berners Lee (traduit en français ici : http://www.urfist.cict.fr/archive/lettres/lettre28/lettre28-22.html) paru en 2001 dans la revue *Scientific American*.

Dans sa vision, Tim Berners Lee cherchait à rendre compte d'un web non plus seulement compréhensible par les humains mais également par les machines. Pour ce faire, les machines devaient "comprendre" c'est-à-dire exploiter la structure des pages web. Or, cette structure est actuellement en HTML et ne permet pas une exploitation en tant que telle. Si les humains savent ce que les balises représentent, quand bien même on les nomme "balises sémantiques" (par abus de langage), elles ne transportent aucune signification (au sens linguistique) et les machines n'ont pas le moyen de déchiffrer de telles structures. Le cœur du web sémantique est donc de "sémantiser" les ressources Web en développant un langage de balises sémantiques qui décrivent les ressources. Ce langage c'est le RDF ou *Ressource Description Framework* (cadre de description de resources), un standard du W3C. Des agents logiciels peuvent ainsi mieux exploiter la masse d'informations, la qualifier, la traiter et donc la rechercher.

**Un cas concret**

Illustrons ce discours par un exemple concret où nous allons comparer des ressources structurées en HTML et en RDF :

*En HTML - Premier exemple de code :*

```
<div class="contenu">
      <p class="normal">
L'article de Myriam Maloron, linguiste informaticienne, qui a pour titre<a href="
http://www.abondance.com/recherche-referencement/">A la découverte du Web Sémantique</a>
décrit les enjeux du Web Sémantique dans la lettre de septembre 2008 sur le site Abondance.
       </p>
</div>
```

*En HTML - Second  exemple de code :*

```
<div class="description">
         <ul>
             <li>
L'article de Myriam Maloron, linguiste informaticienne, qui a pour titre<strong>A la découverte du
Web Sémantique</strong> décrit les enjeux du Web Sémantique dans la lettre de <em>septembre
2008 sur le site Abondance</em>
             </li>
         </ul
</div>
```

Ces deux morceaux de code ne sont pas identiques au niveau de la structure, pourtant ils décrivent et contiennent la même information. Le message est le même et ce que nous indiquent les balises sont la disposition typographique/hiérarchique des divers morceaux  du message, les liens auxquels il renvoie dans le contexte de la page Web, mais ne donne pas d'informations sur ce message. Pour pouvoir décrire ce message (il s'agit d'un article, avec un auteur, sa fonction, un lieu où trouver l'article, etc) il faut sémantiser la structure HTML en y ajoutant du sens. C'est le rôle de RDF...

*En RDF - Premier exemple :*

```
<div class="contenu" instanceof="cal:Vevent">
           <p class="normal">
L'article de <span property="cal:author">Myriam Maloron</span> span
property="cal:job">,linguiste informaticienne,</span><a href="
http://www.abondance.com/recherche-referencement/"><span property="cal:summary"> A la
découverte du Web Sémantique </span></a> décrit les enjeux du Web Sémantique dans la lettre
de septembre 2008
<span property="cal:location" content="20080216T0930-0100> sur le site Abondance </span>.
           </p>
</div>
```

*En RDF - Second exemple :*

```
<div class="description" instanceof="cal:Vevent">
         <ul>
                  <li>
L'article de <span property="cal:author">Myriam Maloron</span> span
property="cal:job">,linguiste informaticienne,</span><a href="
http://www.abondance.com/recherche-referencement/"><span property="cal:summary"> A la
découverte du Web Sémantique </span></a> décrit les enjeux du Web Sémantique dans la lettre
de septembre 2008
<span property="cal:location" content="20080216T0930-0100> sur le site
Abondance</span>.
                  </li>
         </ul>
</div>
```

Dans cet exemple sont décrites les informations suivantes :
- L'article a pour auteur : Myriam Maloron *property="cal:author"*
- L'auteur a une fonction : linguiste informaticienne *property="cal:job"*
- L'article a pour thème : "A la découverte du web sémantique" *property="cal:summary"*
- L'article se trouve sur le site d'Abondance *property="cal:location"*

Certes, à la lecture du message, les informations sont déjà présentes. C'est vrai pour un humain qui lit et comprend ce message. En revanche, ça ne l'est pas pour une machine. L'intérêt du RDF est de rendre possible la "compréhension" de ce message par les machines (applications, robots de search,...) en introduisant du sens dans les balises.

L'intérêt en terme de pratique quotidienne du Web est, cette fois, de rendre exploitables les structures HTML par les moteurs de recherche afin de trouver l'information la plus précise et pertinente en regard d'une requête.

Ainsi, ce bout de code sémantisé permettrait de répondre à la question "Qui a écrit l'article "A la découverte du Web Sémantique" ? Quelle est la fonction de Myriam Maloron ?", etc.

Des outils exploitent déjà les données ainsi structurées, notamment des moteurs de recherche tels que Powerset, Hakia, mais aussi Search Monkey, à des niveaux différents. Nous reviendrons sur ce point lors d'un prochain article de cette série consacrée au Web Sémantique.

### *Qu'en est-il de la démarche de recherche par mots-clés ?*

Le propre du web sémantique est de s'affranchir de la recherche par mots-clés pour aller vers une recherche en langue naturelle. L'internaute requêtera alors en formulant des phrases et non plus des mots-clés.

Cela entraîne une nouvelle pratique du web dans son mode de questionnement mais aussi dans son mode de construction de site.

Une large place sera faite au contenu dans la mesure où c'est lui qui sera structuré et sémantisé. Plus il y aura de contenu informatif et pertinent plus le texte sera à même d'être finement interprétable par les robots de search.

### *Evolution du Web*

La comparaison des différentes étapes dans l'évolution du Web insiste, à terme, sur la mise en relation des données sans passer par l'interprétation humaine. C'est-à-dire sans passer par son interprétation linguistico-cognitive. Ce sont les machines qui interprètent les données.

#### *Différences web traditionnel / web sémantique*

*Intitulé :*
- web informel /web formalisé ;
- web quantitatif /web qualitatif ;
- web anarchique /web informatique ;
- web brouillon /web indexé ;
- web unidimentionnel /web tridimensionnel ;

*Contenus :*
- fait de mots /fait de concepts ;
- compréhensible par les humains /compréhensible par les humains et les logiciels ;
- contenus amorphes /contenus structurés ;
- accumulation /intégration ;
- contenu quelconque /contenu de valeur ;

*Répartition de l'information :*
- concurrence /complémentarité ;
- saturation /synergie ;

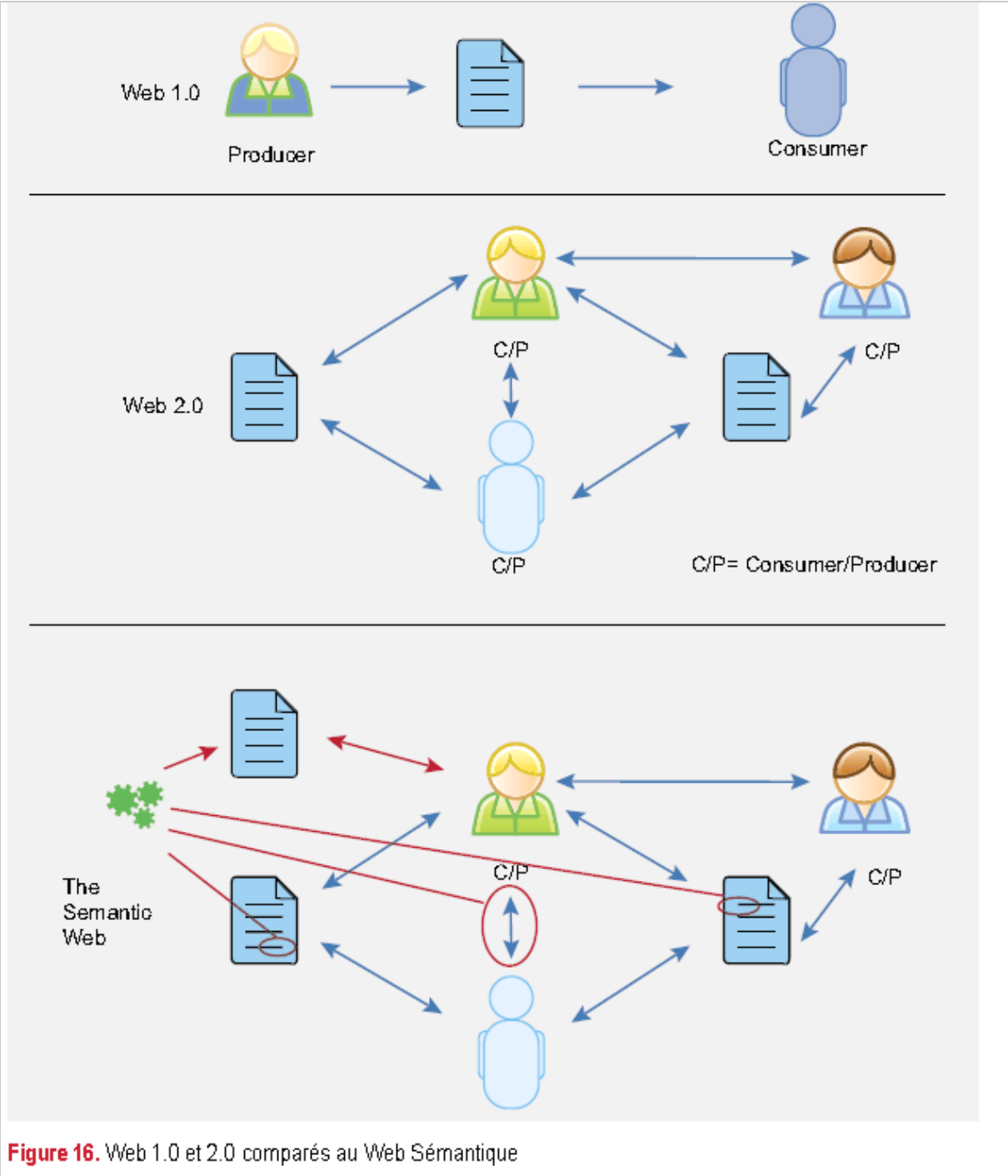


**Figure 16.** Web 1.0 et 2.0 comparés au Web Sémantique

*Comparé au Web 1.0 et 2.0, le web sémantique permet donc de relier les données entre elles" selon leurs affinités", de les faire interagir, et ce, grâce à leur sémantisation.*
*Source : Camille Roux, 2008, www.camilleroux.com*

***Vue sur l'évolution du Web et sur son futur***

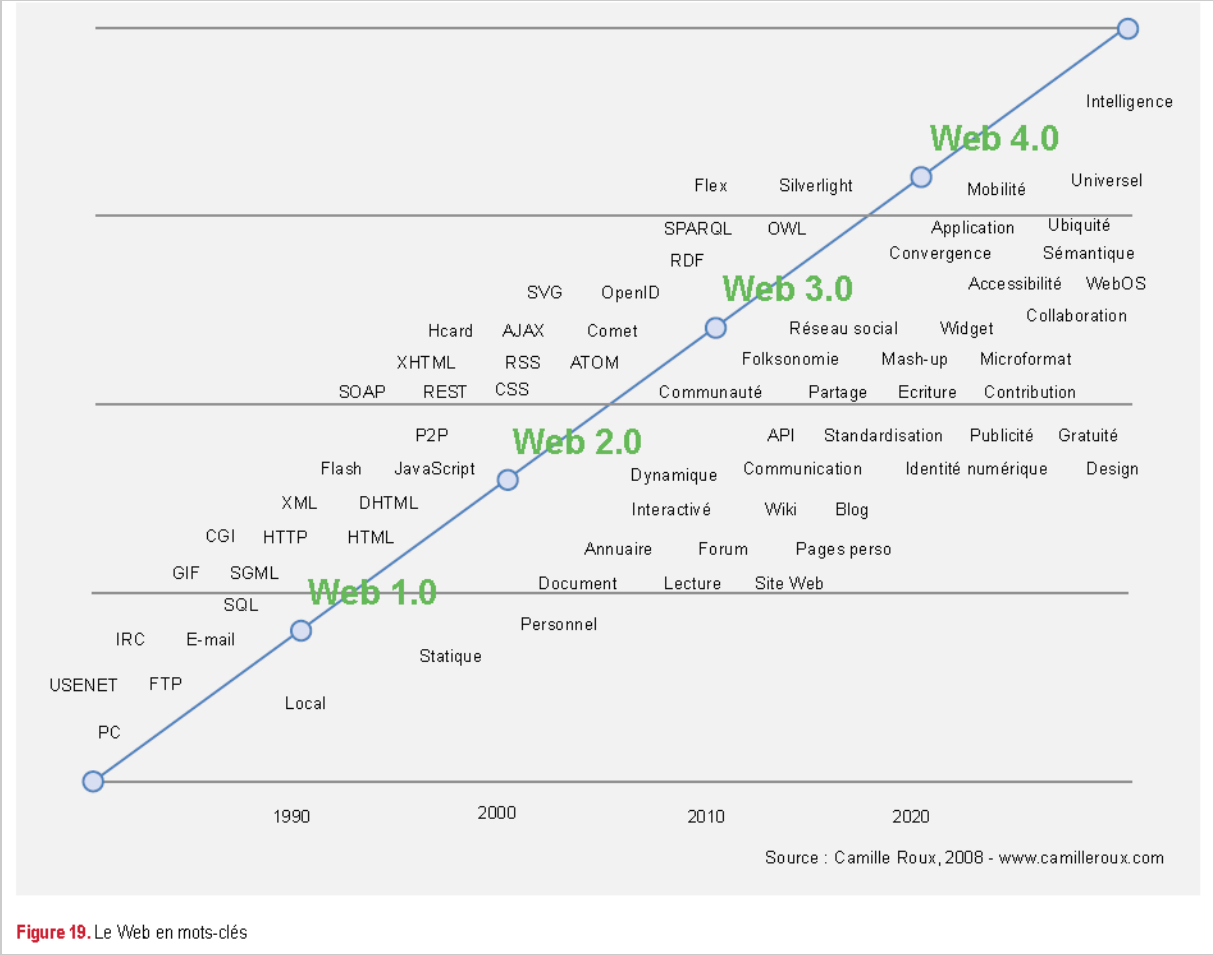


**Figure 19.** Le Web en mots-clés

*Source : Camille Roux, 2008, www.camilleroux.com*

### *Après le Web 3.0...le 4.0 ?*

Ce pourrait être en toute logique la continuité du Web sémantique. Etant donné la base colossale de données désordonnées que représente le Web aujourd'hui, la contribution de robots intelligents qui les structurent et les comprennent semble de bon augure. Le monde du Web 4.0 pourrait bien être celui de l'intelligence artificielle...

***Myriam MALORON***
***veille@brioude-internet.fr***
***Brioude Internet Référencement***

**Réagissez à cet article sur le blog des abonnés d'Abondance :**
**http://abonnes.abondance.com/blogpro/2008/09/la-dcouverte-du-web-smantique-1re.html**