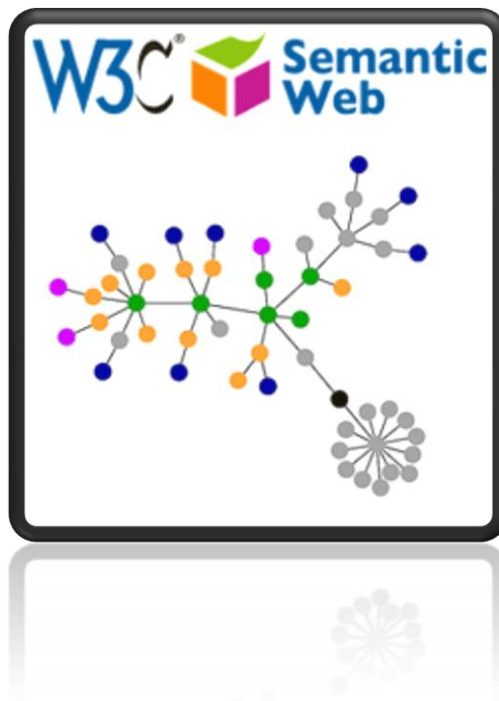


Etat de l'art sur les Technologies du Web sémantique



BOURHIS ANTOINE – PERRIN NASTASSIA

Sommaire

I. ETAT DE L'ART

1. Evolution historique du Web

Le web d'aujourd'hui

Le contexte, l'émergence d'une idée

2. Le web sémantique

Généralités

Architectures et langages

Pré requis : Le Langage XML

Les métadonnées : le socle incontournable

RDF et RDFS: intégrer et interpréter les métadonnées

Les ontologies : formalisation et réseaux sémantiques

Un langage de requête indispensable

Langages de description et de composition de service

II. REFLEXION ANALYTIQUE

1. Comparaison du web 2.0 au web sémantique

2. Moteurs de recherches

Les moteurs de recherche actuels

Le clustering

Moteur de recherche encyclopédique : Wolfram Alpha

Les moteurs de recherche sémantiques

Connaître la qualité d'un moteur de recherche

3. Les applications du web sémantique

L'e-learning

Les Wiki sémantiques

L'intelligence Artificielle

Systèmes experts et moteurs d'inférence

4. Perspectives critiques

Les discours utopiques

Les discours critiques

III. PARTIE PLUS PRAGMATIQUE

1. Moteurs de recherche sémantique

2. True knowledge

3. Cycle de vie du web sémantique

4. Méthodologies et outils

5. OWL vs OWL2

IV. WEBOGRAPHIE

Etat de l'art

1. Evolution historique du web..

Le web d'aujourd'hui

Un peu d'histoire...

L'idée d'internet a débuté en 1945 lorsque Vannevar Bush décrit dans l'article « As we may think » de la revue « The Atlantic Monthly » ce qu'il nomme le Memex (Memory Extender). C'est un appareil électronique qui est relié à une bibliothèque capable d'afficher des livres et de projeter des films.

En 1958 apparaît le premier modem qui permet la transmission de données binaires sur une ligne téléphonique. Puis vient en 1961, la première théorie sur l'utilisation de la communication de paquets pour le transfert de données, théorie développée par L.Kleinrock du « Massachussets Institute of technology ».

Dans le contexte de la guerre froide et sous la menace nucléaire de l'URSS, Us Air Force demande à des chercheurs de créer un réseau de communication capable de résister à une attaque. Ce n'est qu'en 1964 que Paul Baran a l'idée de créer un réseau sous forme de grande toile. Il développe un système centralisé qui n'était cependant pas fiable car toute destruction du noyau provoquait l'arrêt des communications. Par la suite il propose un réseau hybride, d'architectures étoilées et maillées. Les données s'y déplacent dynamiquement en trouvant le chemin le moins encombré et en patientant en cas d'encombrement. C'est le « packets switching ».

Internet prend forme en 1969 lorsque DARPA¹ crée un réseau expérimentale : le premier réseau à transfert de paquets nommé ARPANET. Ce réseau relie quatre universités : Stanford Institut, l'université de Calofornie à Los Angeles, l'université de Californie à Santa Barbara et l'université de l'Utah. L'ARPANET peut être considéré comme l'ancêtre d'internet. La destruction d'un ou plusieurs nœuds du réseau n'entraîne pas la perturbation du fonctionnement général et les communications entre machines se font sans l'intermédiaire d'une machine centralisée.

En 1974 est créé le protocole TCP qui remplace le protocole NCP. Deux ans plus tard ce protocole est déployé sur ARPANET. Puis apparaissent les protocoles TCP/IP en 1978 et DNS en 1984.

Cependant, si internet permettait de faire communiquer toutes sortes de machines, il ne procurait aucun service quant à leur interopérabilité : ce qui était transmis à partir d'un ordinateur était le plus souvent illisible par la machine destinataire.

Travaillant au service du centre de calcul du CERN, Timothy Berners Lee était régulièrement confronté à ce type de problème qui requerrait systématiquement des développements spécifiques.

¹ Defense Advanced Research Projects Agency

D'un laboratoire à l'autre, les physiciens s'échangeaient principalement des textes sur leurs travaux et des images, illustrant leurs théories ou issues de leurs observations expérimentales. Les documents des chercheurs sont par nature émaillés de références bibliographiques et de renvois.

Pour apporter aux chercheurs le service de pouvoir échanger des documents bien présentés, avec des références bibliographiques pouvant donner accès directement aux documents cités et sans souffrir des difficultés inhérentes à la diversité des systèmes informatiques, Tim Berners Lee a proposé et défendu l'idée d'un standard documentaire commun accessible sur tous les systèmes.

À partir de travaux publiés, notamment par IBM, sur des standards documentaires (SGML) d'une part et sur des standards d'hyperliens (Hytime) d'autre part, Berners Lee a donc proposé un modèle spécifique de documents fondés sur le standard SGML et incorporant les fonctionnalités de navigation via des liens hypertextes.

En 1989, Berners-Lee propose l'idée de naviguer simplement d'un espace d'internet à un autre, à l'aide de liens hypertextes grâce à un navigateur. Cette proposition a pour but d'améliorer la transmission d'information au sein même du CERN et avec l'ensemble des laboratoires de recherche en physique nucléaire et en astronomie.

L'ingénieur Robert Cailliau rejoint le projet et en fait sa promotion. Le premier prototype fut *WorldWideWeb* (plus tard il sera rebaptisé *nexus*). Ces deux personnes sont à l'origine du WWW².

Le premier système WWW, mis à la disposition des physiciens via la bibliothèque de logiciels du CERN, fut créé en 1991. Muni d'un navigateur simple, d'un logiciel de serveur web et d'une bibliothèque exécutant les fonctions essentielles, ce système permettait aux développeurs de bâtir leur propre logiciel. Il fut utilisé par le milieu universitaire et en laboratoire avant d'être mis en libre accès.

L'équipe du CERN ne pouvant pas à elle seule faire avancer le système, Tim Berners-Lee lance un appel sur internet afin de trouver des développeurs capables d'apporter leurs compétences.

La première avancée sera en 1993 lorsque l'université de l'Illinois présente son navigateur web graphique : Mosaic. C'est le début des interfaces graphiques des navigateurs qui entraînent une croissance de la popularité du web. La Netscape Communications Corporation propose alors en 1994 le navigateur Netscape Navigator qui devient le navigateur le plus populaire. Microsoft entre ensuite dans la course aux navigateurs avec Internet Explorer en 1995 qui surpasse Netscape auprès des utilisateurs. Plus tard, Netscape lance le projet Mozilla qui développe Firefox.

En 1994, Berners-Lee fonde le World Wide Web Consortium, appelé W3C, dont l'objectif est d'émettre des recommandations afin de promouvoir et d'assurer la compatibilité des technologies utilisées sur le Web. Tim Berners-Lee en est toujours le directeur aujourd'hui.

² World Wide Web

Le web 1.0 et le web 1.5

Le web initial (web 1.0 ou web statique) était initialement confiné à la communauté des chercheurs. Il s'est répandu auprès du grand public via la publication de pages HTML alliant du texte, des liens et des images, et il est consultable en ligne via un navigateur web grâce au protocole http. Ce sont les entreprises (propriétaires de ce qu'on appelle un site) qui produisent et hébergent le contenu. Par manque d'outils appropriés, ces sites sont rarement mis à jour. Avec le web 1.0 l'utilisateur n'est alors qu'un simple lecteur de l'information mise à disposition.

De nouveaux langages de scripts apparaissent (tels que les langages Php, javascript) qui, utilisés avec des bases de données, permettent de rendre dynamique la publication des pages sur les sites. C'est la migration vers le web 1.5. Des personnes sont alors capables, grâce à ces scripts, de gérer un site. Ils peuvent ainsi créer des pages ou des articles voire insérer des module du type e-mailing-list, flux RSS, heure et date, etc. Puis sont apparus les forum, espaces d'échanges par le biais de messages.

Microsoft développe en 1998 l'objet XMLHttpRequest pour Internet Explorer 5.0 qui permet d'obtenir des données au format xml, json, html ou texte via des requêtes http. Cette avancée marque le début du web 2.0. L'objet XMLHttpRequest n'est plus supporté par les anciens navigateurs, il est donc implémenté par la suite sous Mozilla 1.0 (mai 2002), Safari 1.2 (février 2004), Konqueror 3.4 (mars 2005) puis Opera 8.0 (avril 2005).

Le web 2.0

Le web 2.0, appelé aussi web participatif ou web collaboratif, permet la visibilité de l'information à tous, et à n'importe quel internaute d'être actif car les contenus sont produits et réalisés par les internautes. Il s'agit moins d'une évolution technique que d'un phénomène sociologique, en quelque sorte "libéré" par la possibilité donnée aux internautes d'intégrer avec les sites via Ajax et XMLHttpRequest.

Il existe de nombreuses définitions du web 2.0. Pour certains il s'agit d'un mélange de sites ou d'outils qui encouragent la collaboration et la participation. Les visiteurs peuvent y partager leurs photos, leurs vidéos, leur vie ou leurs passions. Apparaissent ainsi les réseaux sociaux qui permettent de définir d'autres utilisateurs comme étant leurs amis, leurs collègues, leurs contacts donc de créer des relations entre les utilisateurs. Une seconde définition présente le web comme le fait de gagner de l'argent en finançant un site dont le contenu est alimenté par des utilisateurs/internautes.

On peut aussi retenir la définition faite par Tristan Nitot, président de Mozilla Europe, lors d'une interview du Journal du Net³ : « *Encore faudrait-il comprendre ce qu'est le Web 2.0 : tout le monde a sa propre définition de la chose... Pour moi, le Web 2.0, c'est le Web tel qu'il a été imaginé par son inventeur, Tim Berners-Lee, un Web où chacun peut publier et consommer de l'information. Un Web où l'on est consommateur, et acteur à la fois. Consomm'acteur, en quelque sorte.* »

Cependant Tim Berners Lee devait proposer en 2006 une autre vision, d'un "Web 3.0" : *"People keep asking what Web 3.0 is. I think maybe when you've got an overlay of scalable vector graphics - everything rippling and folding and looking misty — on Web 2.0 and access*

³ <http://www.journaldunet.com/>

to a semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource."

Le contexte, l'émergence d'une idée

Tim Berners Lee avait introduit la notion de Web sémantique huit années plus tôt. En 1998 dans deux articles « Semantic Web Road Map » et « Weaving the web », il présente pour la première fois le web sémantique sans préciser sa vision, de peur que tout le monde se jette sur l'idée et gache le principe. C'est cette même année qu'est émise la première recommandation XML, modèle documentaire fermé sur un jeu défini de balises. Il était nécessaire d'ouvrir le concept, de laisser libre la définition du balisage et de ses attributs pour permettre d'échanger d'autres expressions que des documents destinés aux navigateurs, à la façon de ce que permettait le langage SGML dont HTML est un modèle (DTD) particulier. Cependant les documents SGML avaient l'inconvénient de ne pouvoir être lus sans connaissance de leur modèle, ce qui nécessitait des outils coûteux à développer et à utiliser. XML propose comme SGML un format textuel d'échange de données structurées.

Il faudra attendre mai 2001 pour que Tim Berners Lee développe le concept du web sémantique à la 10^e conférence du World Wide Web à Hong Kong. Il annonce officiellement la première recommandation pour les schémas XML, permettant d'unifier l'expression des modèles XML en XML soit même, en remplacement du langage de DTDs hérité de SGML.

Le même mois, Berners Lee, James Hendler (chargé de recherche sur l'informatique) et Ora Lassila (chercheur chez Nokia et membre du conseil de direction du Consortium W3) publient l'article « The Semantic Web » de la revue *Scientific American*. Cet article commence par un scénario qui se passe dans un futur proche dans lequel Pete et sa sœur Lucy trouvent une solution rapide à leur problème grâce au web sémantique. Voici un résumé de ce scénario :

« Lorsque Lucy appelle chez Pete, tous ses appareils sonores baissent automatiquement le volume. Son téléphone ajuste le son en envoyant un message aux différents appareils possédant un contrôle du volume. Pete et Lucy ont besoin de conduire leur mère dans un hôpital qui pourra lui dispenser les soins dont elle a besoin. Ils font de suite une recherche avec leurs logiciels agents, ce qui permettra de leur trouver un hôpital qui répondra à leurs critères : proche, qui dispense les soins recherchés, et dont les horaires sont compatibles avec les leurs. En proposant cette solution, le système émet deux avertissements : le fait que Pete aura à annuler deux de ses rendez-vous et des détails sur l'assurance de leur mère. »

Cet exemple nous montre bien que pour Berner Lee, le web sémantique dépasse le cadre des réseaux et de l'informatique. Il est destiné à faire partie intégrante de notre vie quotidienne et permettre de faire interagir des objets de différentes natures.

Nous allons donc expliciter ce concept.

2. Le web sémantique

Généralités

Le web sémantique constitue un environnement dans lequel les humains et les machines vont communiquer selon une base sémantique. La première définition qui ressort lorsque l'on parle de web sémantique est celle d'un web compréhensible par les machines. Il s'agit donc d'un concept à la fois large et complexe définissable de la façon suivante : ensemble de technologies visant à rendre le contenu des ressources du web accessible et utilisable par les programmes, agents logiciels et machines, grâce à un système de métadonnées formelles (données servant à décrire d'autres données), utilisant notamment la famille de langages développés par le World Wide web Consortium (W3C).

De nos jours, les volumes d'information sur le web augmentent considérablement, ce qui renforce le besoin de partager, d'échanger et de réutiliser ces informations d'une manière efficace entre les utilisateurs. La plupart des informations se trouvant sur Internet sont effectivement lisibles par les hommes mais souvent difficilement interprétables par les machines. En outre, les moteurs de recherche utilisant les mots clés ne tiennent pas compte de la sémantique qui les lie entre eux. Ainsi, une recherche de plusieurs mots ne donne pas nécessairement les bons résultats car le moteur ne tient pas compte de la sémantique associée à l'ensemble de ces mots, c'est-à-dire du sens de la phrase dans sa globalité. Les pages retournées par le moteur de recherche peuvent contenir les mots clés mais ne sont pas obligatoirement les plus pertinentes, elles ne tiennent pas compte de la sémantique associée aux mots clés reliés entre eux. La recherche peut alors être dans certains cas imprécise. On pourrait donc se demander pour quelles raisons cela arrive. Pourquoi un moteur de recherche basé sur des mots clés ne détient pas forcément la sémantique générale associée à la requête utilisateur ? La raison principale est que la description des informations (les métadonnées) est limitée, ou alors non exploitée d'une manière plus "intelligente" par les machines.

Aujourd'hui, le manque de normes et de sémantique formelle des informations disponibles sur le web empêche également le partage optimisé d'information entre les individus et les organisations. Par conséquent, la vision du web du futur a pour objectif de relier des connaissances en permettant à des machines d'accéder aux sources d'information et aux services du web. Tout repose donc sur un enjeu majeur : décrire correctement les informations disponibles sur le net afin de les rendre explicites facilement par les machines, en particulier au niveau de la sémantique qui relie ces informations entre elles.

Le concept de web sémantique selon Tim Berners-Lee (fondateur du World Wide web Consortium) fait donc d'abord référence à la vision d'un web optimisé représentant le plus large espace d'échange de ressources et supportant intuitivement les relations entre les humains et les machines, dans le but d'exploiter plus efficacement de grands volumes d'informations et de services variés disponibles sur le web. La finalité du web sémantique serait donc de rendre explicite le contenu sémantique des ressources dans le web (documents, pages web, services, etc..) par les machines et agent logiciels, afin qu'ils puissent comprendre le sens des informations via les descriptions contenues dans ces ressources.

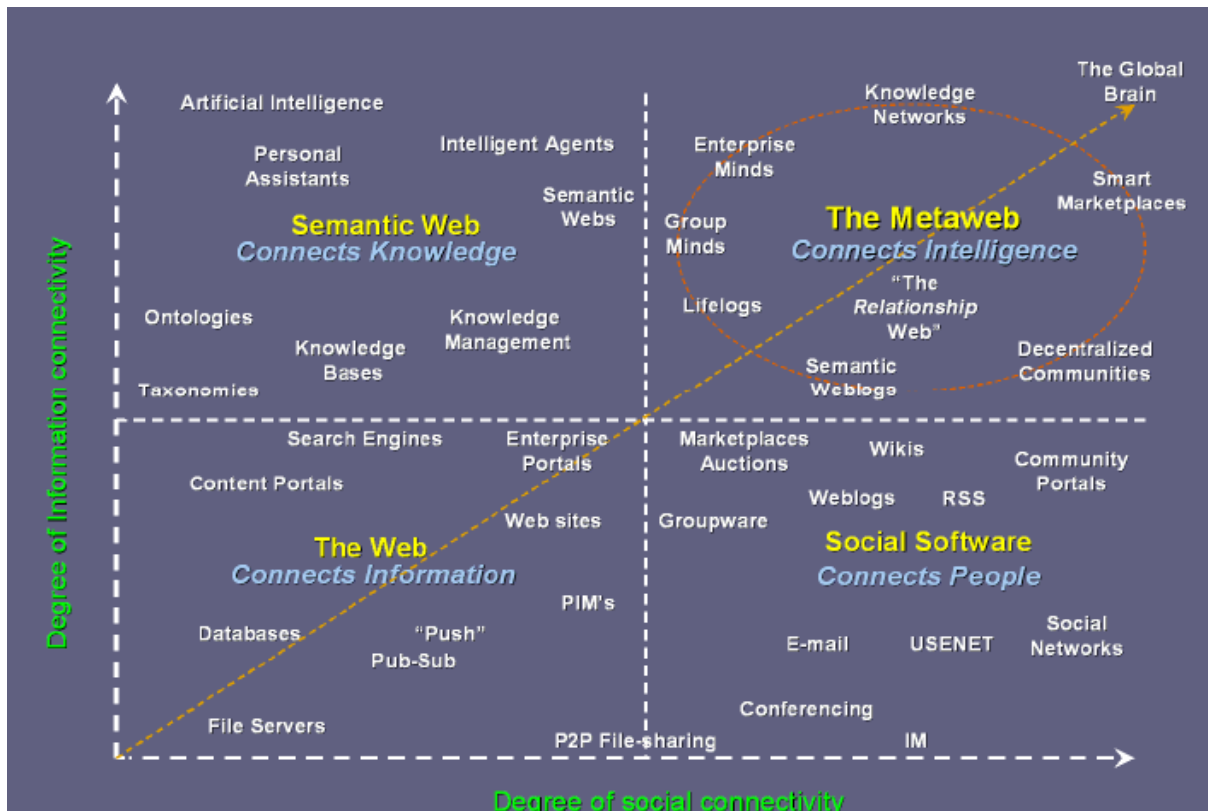


Figure 1 - Le web du futur [Spivack, 2004]

Le schéma ci-dessus montre le processus chronologique destiné au web : l'idéal dans le futur serait de pouvoir associer la notion de web sémantique avec celle des applications sociales, dans le but de créer un « Metaweb » focalisé sur le partage d'informations entre communautés.

Les recherches réalisées sur web sémantique s'appuient sur un existant provenant d'environnements très variés mais intrinsèquement liés par des concepts inhérents au web sémantique, telle que la philosophie, l'informatique et la linguistique. En effet, on trouve des applications concrètes de web sémantique dans les recherches intelligentes d'information, sur l'ingénierie des connaissances, les systèmes de représentation des connaissances, l'apprentissage, le traitement automatique de la langue naturelle ou encore le raisonnement automatique...

Dans cette partie, nous allons donc présenter en détail la démarche du web sémantique au travers de l'architecture et de certains langages de représentations qui sont largement utilisés pour celui-ci.

La vision courante du web sémantique proposée par Berner Lee est représentée dans une architecture en plusieurs couches différentes que nous allons expliciter (cf. Figure ci dessous).

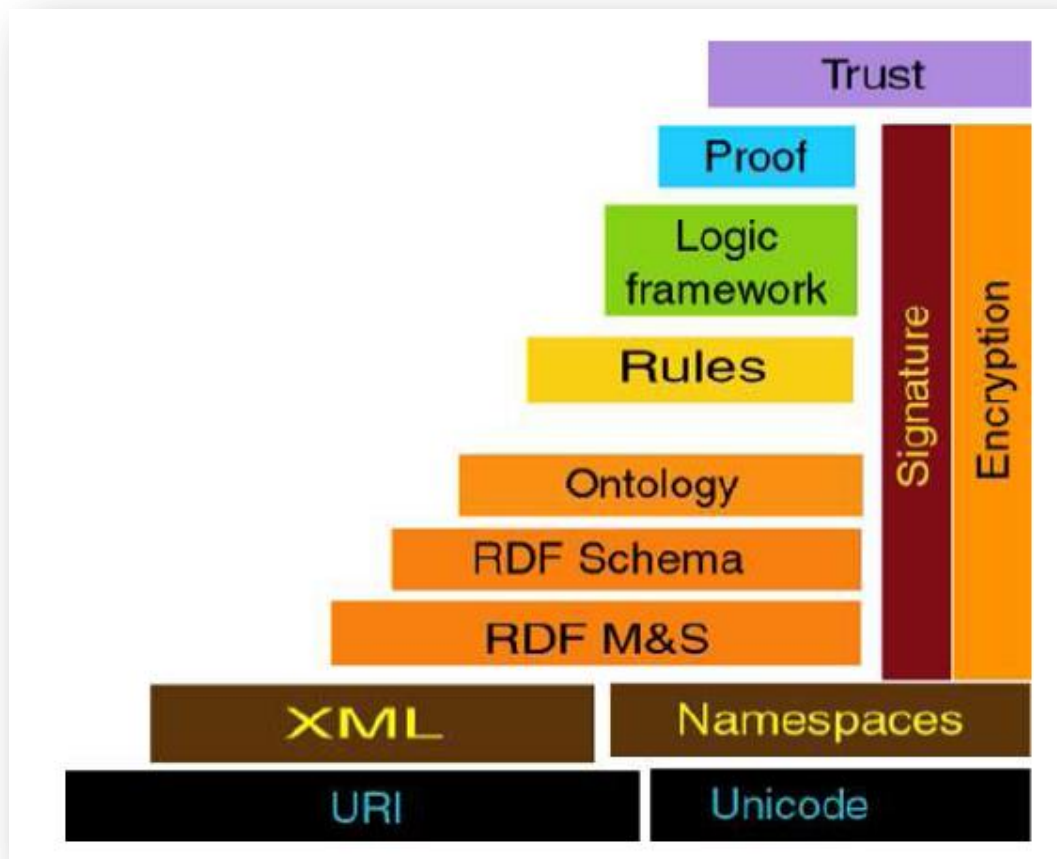


Figure 2 : Architecture pyramidale du web sémantique [W3C]

- Les couches les plus basses assurent l'interopérabilité syntaxique :
 - La notion d'URI (Uniform Resource Identifier, identifiant uniforme de ressource) fournit un adressage standard universel permettant d'identifier des ressources : c'est une courte chaîne qui fournit l'identification de la ressource sur un réseau.
 - Unicode est un encodage textuel universel pour échanger des symboles.
- XML est un langage proposant une syntaxe de description de la structure d'un document, mais aussi de création et de manipulation de documents. Il utilise un espace de nommage (namespace) permettant d'identifier des balises utilisées dans les documents Xml. Le document est validé et définit par un schéma. Cependant, XML n'impose aucune contrainte sémantique à la signification de ces documents. Xml ne suffit donc pas pour assurer aux logiciels la compréhension du contenu des données incluses dans le document : l'interopérabilité syntaxique n'est pas suffisante.
- Les couches RDF M&S (RDF Model & Syntax) et RDF Schéma sont considérées comme les premières fondations de l'interopérabilité sémantique. Leur rôle essentiel est de décrire des lois de classification de concepts ou de propriétés sur des données.
 - RDF est considéré comme une fondation car c'est le premier langage fournissant un moyen d'insérer de la sémantique dans un document.

- Le schéma RDFS décrit les hiérarchies des concepts et des relations entre les concepts.

- La couche ontologie apporte une évolution car elle assure la description de sources d'information hétérogènes. Ces sources peuvent d'ailleurs formaliser une conceptualisation de choses existantes partagée par plusieurs personnes, voir par toute une communauté. Le rôle de l'ontologie est donc d'aider l'humain et la machine à communiquer, en priorisant sur l'échange de sémantique des informations plutôt que la syntaxe et en utilisant des règles précises.
- La couche Rules (règles) offre les moyens de l'intégration, de la dérivation, et de la transformation de données provenant de sources multiples.
- La couche Logique se trouve au-dessus de la couche Ontologie. Certains les considèrent comme étant au même niveau.
- Les couches Proof (Preuve) et Trust (Confiance) permettent de vérifier des déclarations effectuées dans le web sémantique.

Le schéma ci-dessous montre de manière plus détaillée les étapes du web sémantique : on retrouve les paliers de la pyramide et la vision stratégique énoncée par Berner Lee : la démarche est détaillée ici de façon plus explicite.

Si l'on part du web sémantique, qui est le but en soit, on peut voir qu'il dépend des différents agents (logiciels ou machines), qui eux même dépendent d'un service de requête (Query Service dépendant du langage XML), mais aussi de la couche confiance (trust), elle-même dépendant des couches Preuve (Proof) Sécurité (Security) et Règles associées au langage de requête (Rules). Les agents dépendent également de cette couche Règles, qui est basée sur la définition des ontologies, elle-même existant par le biais des seules métadonnées. On retrouve donc ici clairement la démarche stratégique associée au web sémantique.

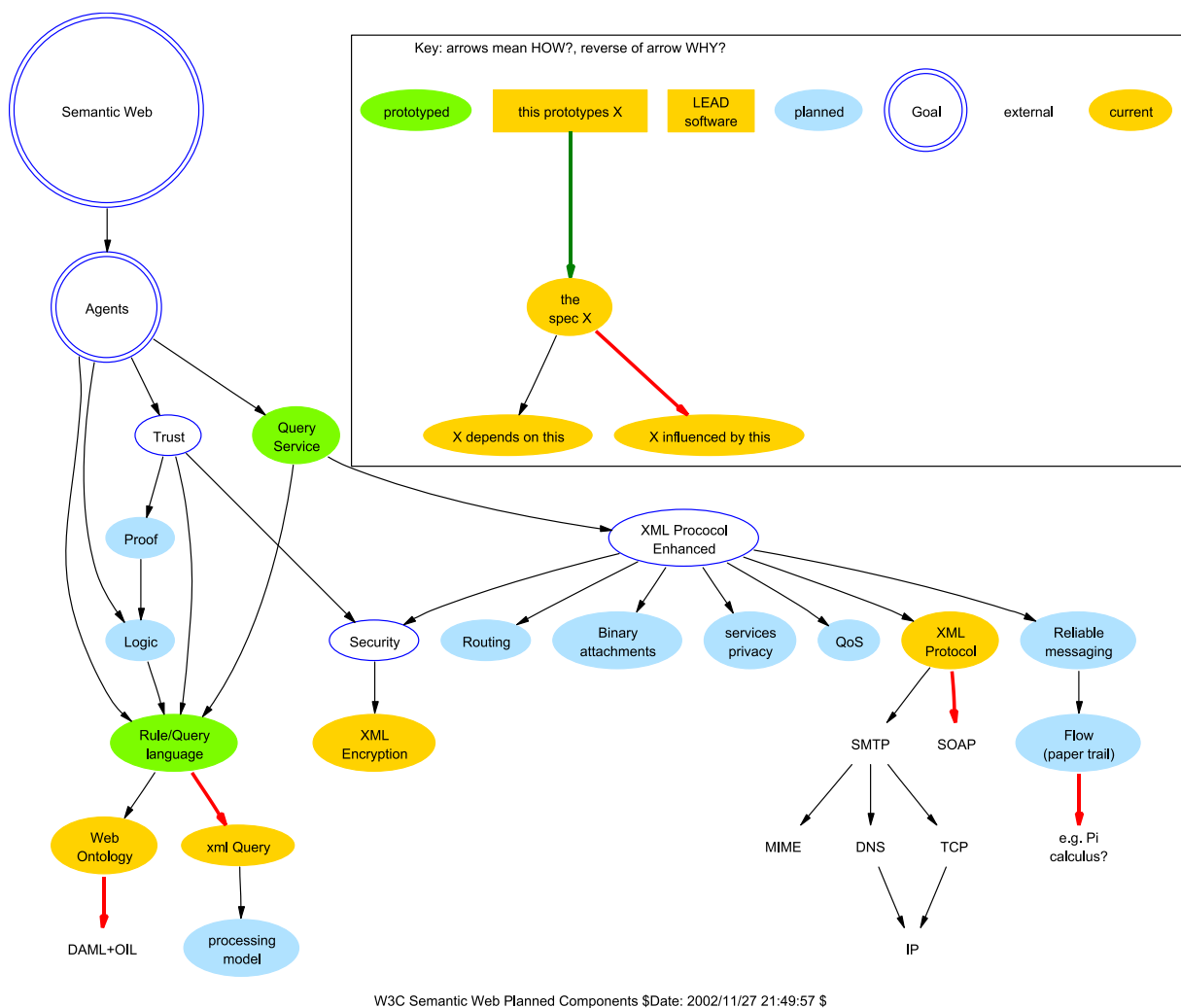


Figure 3 : Semantic Web Roadmap [W3C]

Architectures et langages

Dans cette partie, notre démarche va consister à expliquer de manière détaillée les différents concepts inhérents au web sémantique, c'est-à-dire les différents piliers qui lui servent de socle. Une fois ces concepts détaillés, il conviendra de décrire les différents moyens d'implémentation ainsi que les outils utilisés pour cela, dont le but est d'améliorer la communication et l'interopérabilité des applications du web sémantique.

Comme nous l'avons expliqué dans la présentation, la démarche du web sémantique est ancrée dans une volonté de rendre le web accessible à tous, agents logiciels, humains, machines, et ce de façon automatique, industrielle, unifiée et pérenne. Le but ultime est de mettre en place un web idéal, où tout le monde, tout objet machine ou logiciel peut trouver les informations qu'il désire en focalisant ses requêtes sur la sémantique des contenus web. On peut donc légitimement se demander quelles sont les solutions techniques, langages, architectures et outils qui ont été choisis, qui seront utilisés, et dans quel but...

La base de cette démarche est de proposer un ensemble de métadonnées au sein même des contenus web. Ces métadonnées, ou annotations, vont permettre de décrire, en suivant une syntaxe bien précise, les contenus et les données de la page web. Cette évolution technologique pourrait donc permettre de grandes avancées, à charge au développeur de la page web d'implémenter correctement sa page et de lui fournir les métadonnées suffisantes pour qu'elle puisse être comprise et utilisée par les agents externes. Cette compréhension à partir de métadonnées va être rendue possible grâce à la mise en place de réseaux sémantiques, et l'utilisation de ce que l'on appelle une ontologie.

Pré requis : Le Langage XML

XML est un langage permettant de structurer des informations dans un document. C'est un langage de balisage extensible qui est considéré comme une spécification pour les documents lisible par les machines. C'est le langage de base utilisé par l'encodage des langages du web sémantique. Le balisage signifie que certaines suites des caractères du document peuvent contenir de l'information indiquant le rôle du contenu du document, c'est-à-dire que XML se sert de balises (par exemple <nom> ou <adresse>) pour décrire un classement ou une structure logique des données d'un document. Le caractère extensible indique la différence importante avec d'autres langages précédents qui est aussi la caractéristique essentielle de XML, car XML formalise la structuration d'un document. C'est un métalangage qui fournit une structure pour représenter d'autres langages d'une manière normalisée. Un document XML a donc toujours un élément racine, et des éléments XML peuvent contenir d'autres éléments qui sont représentés par les balises d'ouverture et de clôture. Le web sémantique est très largement tributaire de ce langage, reconnu comme un standard de structuration et d'échange de documents Web. Comme nous l'avons vu dans la pyramide symbolisant la stratégie du web sémantique selon Berner Lee, XML est un pré-requis indispensable au web sémantique.

Exemple d'un morceau du document XML :

```
<?xml version="1.0"?>
<etudiant>
  <nom>BOURHIS</nom>
  <prenom>Antoine</prenom>
  <phone>0650747719</phone>
  <bureau>
    <numero>AA23</numero>
    <batiment>Batiment A</batiment>
  </bureau>
</etudiant>
```

Les métadonnées : le socle incontournable

Il est légitime de se demander comment ces métadonnées ou annotations vont être ajoutées au contenu d'une page web, en suivant des standards et une syntaxe spécifique. Avant cela, il convient de définir cette notion et en quoi elle est essentielle pour le web sémantique. Aujourd'hui, l'information sur le web a tendance à être de plus en plus volumineuse, hétérogène, et distribuée souvent de façon peu structurée. L'utilisation des ressources du web est donc nécessairement tributaire de méthodes et d'outils pour les représenter, les manipuler et les exploiter.

Dans le web sémantique, les métadonnées et annotations sont utilisées au sein des contenus. L'annotation est une information graphique ou textuelle rattachée au document ou placée dans celui-ci. Ces annotations sont souvent des annotations en langage naturel, ou des annotations symboliques, des images et des sons. Elles sont considérées comme des métadonnées lorsqu'elles sont associées à l'objet qui a été annoté. La métadonnée est une donnée relative à d'autre donnée, elle sert de description sur une donnée et est interprétable par des machines. Il y a donc ici une légère subtilité entre ce qui est considéré comme une métadonnée et une annotation : l'annotation représente une nouvelle donnée attachée à une ressource, alors que la métadonnée est une donnée sur une donnée. Une métadonnée sera donc plutôt rattachée à une ressource identifiée en tant que telle sur la toile, en ayant à priori de la pertinence sur la donnée qu'elle décrit, tandis que l'annotation pourra être considérée comme une note critique ou explicative accompagnant une donnée. Néanmoins ces deux notions sont très largement comparables dans le contexte du web sémantique.

Un des objectifs dans l'environnement du web Sémantique est de décrire le contenu des ressources en les annotant avec des informations non ambiguës, et ce pour favoriser l'exploitation de ces ressources par les agents logiciels ou machines.

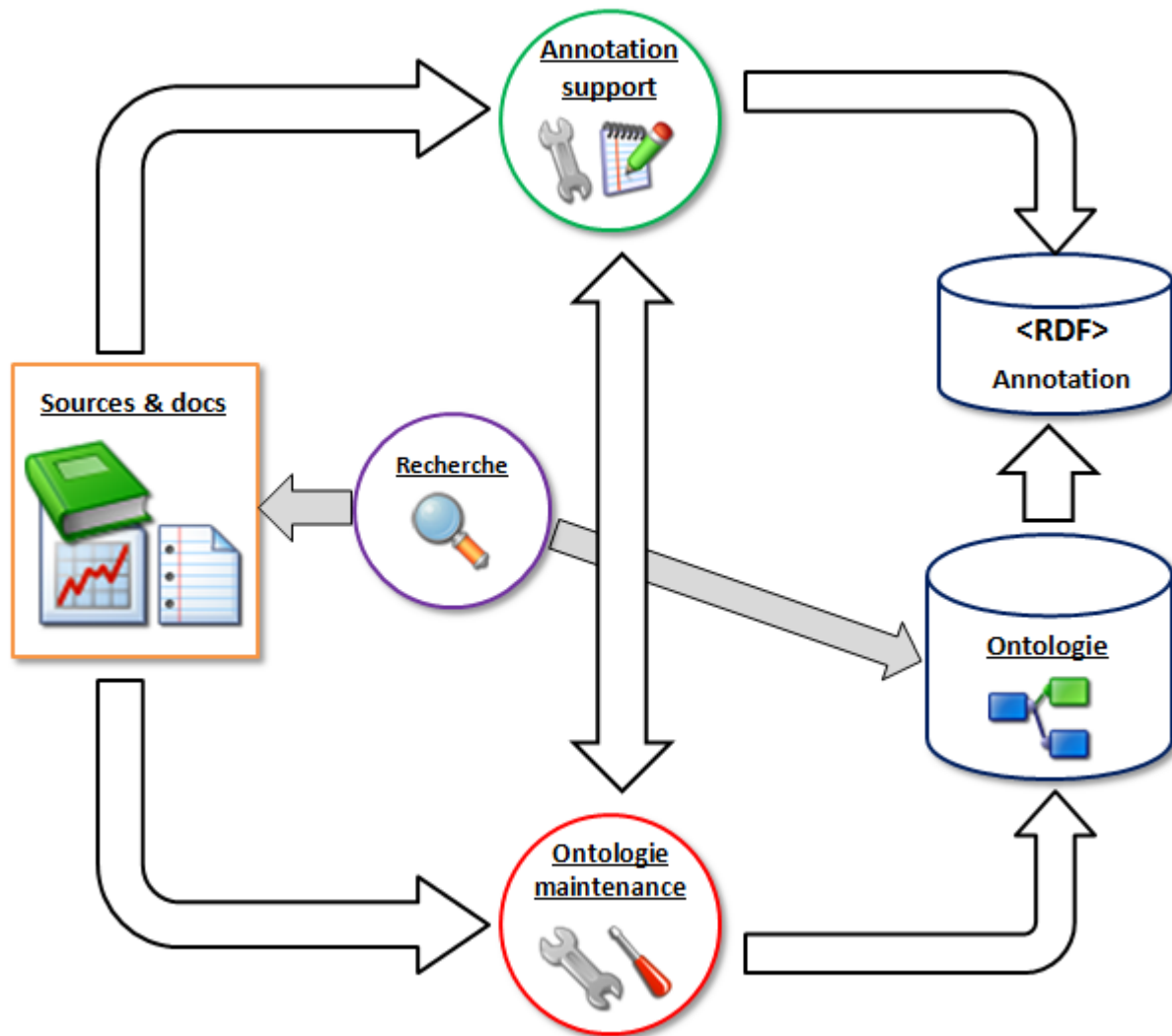


Figure 4 : Rôle des métadonnées dans un système de gestion de connaissances

La Figure ci-dessus représente le rôle de l'annotation dans un système de gestion des connaissances. Les annotations, créées et manipulées par les outils d'annotation, fournissent l'interopérabilité entre différents types de documents et supportent des services de recherche. Les outils de recherche sémantique sont utilisés pour connecter et exploiter les informations attachées dans les annotations et les documents. La recherche est focalisée sur les documents sources ainsi que sur les réseaux ontologiques et annotations qui leur sont associées. Les outils de maintenance de l'ontologie supportent le processus d'évolution et de maintenance des modèles de connaissance pour répondre aux changements. Dans le cadre du web sémantique, les annotations sémantiques ont donc pour objectif d'exprimer la sémantique du contenu d'une ressource.

RDF et RDFS: intégrer et interpréter les métadonnées.

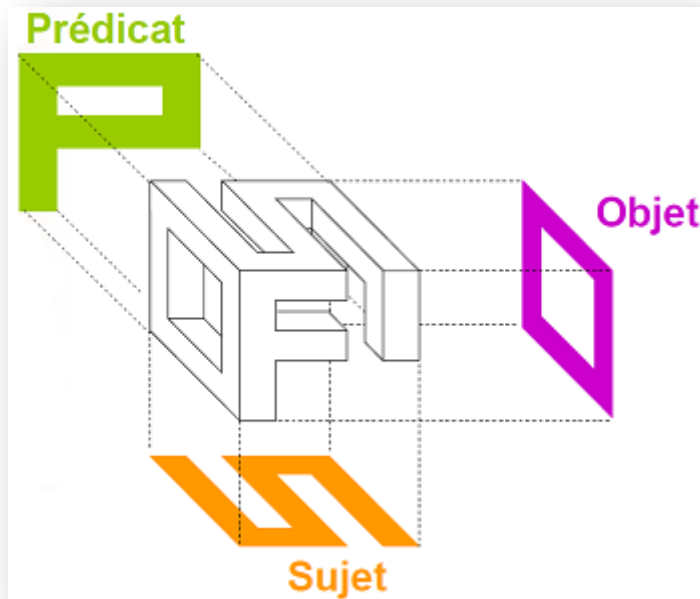
Les langages d'annotation qui vont être utilisés pour comprendre et adapter les différentes annotations des contenus web vont être la première étape importante dans le processus de compréhension des données par les agents logiciels.

RDF (*Resource Description Framework*) est un modèle de graphe destiné à décrire de façon formelle les ressources web et leurs métadonnées. Ce formalisme pour la description des métadonnées, propulsé par le W3C, est le langage de base du web sémantique dans sa forme de compréhension automatique et facile des contenus web. C'est la première couche réelle des standards du web sémantique car ce langage fournit de façon concrète l'interopérabilité entre les applications qui échangent des informations sur le web. Il permet de formaliser les métadonnées dans les documents et surtout de les rendre compréhensibles par le biais d'une formalisation descriptive formalisée et articulée autour de triplets.

La principale syntaxe utilisée par ce langage est RDF/XML, de ce fait RDF est syntaxiquement plutôt proche de XML. Ses objectifs principaux peuvent être synthétisés de la façon suivante :

- RDF permet de manipuler et classer les métadonnées des documents web et de fournir des informations sur les ressources du web et sur les systèmes qui utilisent ces ressources.
- RDF permet également de développer des modèles d'information non cloisonnés et non fixés à certaines applications spécifiques.
- RDF permet aux informations d'être manipulées en dehors de l'environnement spécifique dans lequel elles avaient été créées. C'est le rôle d'interopérabilité des savoirs.
- RDF optimise la coopération entre applications web en permettant de combiner les données de plusieurs applications dans le but de générer des informations nouvelles.
- Enfin, et ce rôle est inclus dans chacun des précédents points, RDF facilite le traitement automatisé des informations du web par les agents logiciels, dans le but de transformer un regroupement d'informations initialement destinées aux humains en un état de réseau de processus de coopération entre eux : RDF joue le rôle de langue privilégiée et compréhensible par tous les agents sur la toile.

Concernant les triplets, ils sont composés d'une part de la ressource concernée (Page, image, vidéo, ou tout autre objet web qui peut avoir un *URI*), mais surtout d'une description de cette ressource (attributs, annotations ou métadonnée...), et enfin d'un framework associé (modèle, langages et syntaxes pour ces descriptions). L'utilité principale de RDF est de permettre le traitement automatique de descriptions des données, c'est-à-dire l'assimilation automatique des métadonnées contenues dans les documents web.



Les ensembles de triplets sont représentables par le biais d'un graphe, et plus spécifiquement un multi-graphe orienté et étiqueté où les éléments apparaissant comme ressource ou valeur sont les sommets. De plus, chaque triplet est représenté par un arc dont l'origine est sa ressource et la destination sa valeur comme la montre la figure suivante.

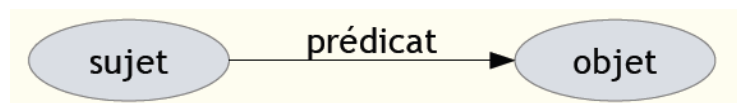
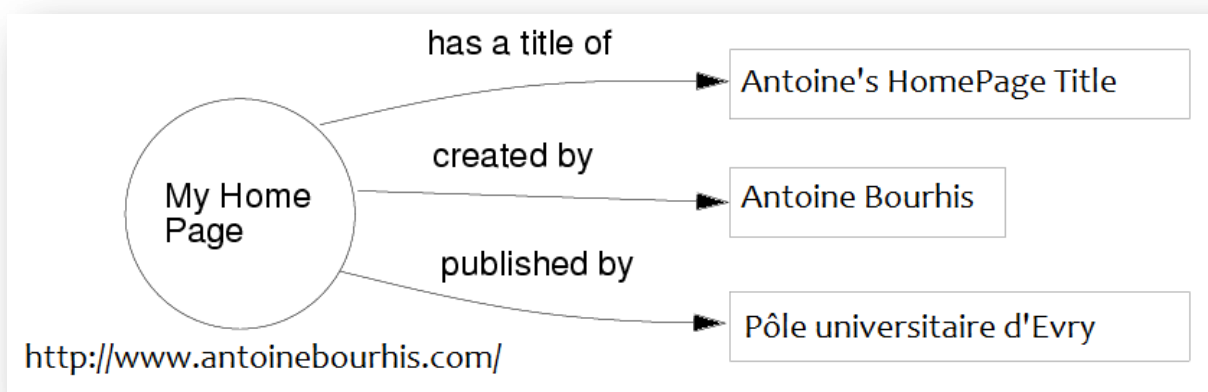


Figure 3. Ensemble de Triplet RDF

Afin d'expliciter plus concrètement le rôle de RDF, voici l'application concrète d'une situation simple (Schéma RDF & implémentation XML)

Ici, ce schéma montre bien les trois états (Sujet/Predicat/Objet) respectivement :



My Home Page, has a title of, "Antoine's HomePage Title"

My Home Page, created by, "Antoine Bourhis"

My Home Page, published by, "Pôle universitaire d'Evry"

Les URI de ressource et prédicats sont les suivants :

```
http://www.antoinebourhis.com, [dc:title], "Antoine's HomePage Title"  
http://www.antoinebourhis.com, [dc:creator], "Antoine Bourhis"  
http://www.antoinebourhis.com, [dc:publisher], "Pôle Universitaire d'Evry"
```

La syntaxe XML pour notre exemple est la suivante :

```
<?xml version="1.0"?>  
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/">  
  <rdf:Description rdf:about="http://www.antoinebourhis.com">  
    <dc:title>Antoine's HomePage Title</dc:title>  
    <dc:creator>Antoine Bourhis</dc:creator>  
    <dc:publisher>Pôle universitaire d'Evry</dc:publisher>  
  </rdf:Description>  
</rdf:RDF>
```

Il existe une sorte d'extension à RDF qui a pour principal but d'étendre le langage RDF en décrivant de manière plus précise les ressources utilisées pour étiqueter les graphes. Cette extension, nommée RDFS pour RDF Schéma, permet de mettre en exergue certaines limites de RDF : le modèle des données RDF ne précise que le mode de description des données mais ne fournit pas la déclaration des propriétés spécifiques au domaine ni la manière de définir ces propriétés avec d'autres ressources. RDF Schéma permet donc de palier à cela en fournissant un mécanisme permettant de spécifier les classes *C* dont les ressources étiquetées par ces classes *C* seront des instances, comme les propriétés.

RDFS s'écrit toujours à l'aide de triplets RDF, en définissant la sémantique de nouveaux mots-clés. Voici un exemple de syntaxe utilisée en RDFS :

```
<ex:Vehicule rdf:type rdfs:Class>
```

Ici la ressource *ex:Vehicule* a pour type *rdfs:Class*, et est donc une classe;

```
<snCF:TER7089 rdf:type ex:Vehicule>
```

Ici la ressource *snCF:TER7089* est une instance de la classe *ex:Vehicule* que nous avons définie précédemment

```
<snCF:Train rdfs:subClassOf ex:Vehicule>
```

Ici la classe *snCF:Train* est une sous classe de *ex:Vehicule*, toutes les instances de *snCF:Train* sont donc des instances de *ex:Vehicule*;

```
<ex:localisation rdf:type rdfs:Property>
```

affirme que *ex:localisation* est une propriété (une ressource utilisable pour étiqueter les arcs).

```
<ex:localisation rdfs:range ex:Ville>
```

affirme que toute ressource utilisée comme extrémité d'un arc étiqueté par *ex:localisation* sera une instance de la classe *ex:Ville*.

RDF et RDFS sont donc les principaux langages d'annotation, permettant d'exploiter de façon simple et rapide les métadonnées contenues dans les documents. Néanmoins, il existe d'autres langages d'assertion et d'annotation qui ont un rôle plus ou moins proche. Nous présentons ici un de ces langages, les cartes topiques (« topic maps »), qui sont un standard issu de la norme ISO et dont le but est d'annoter les documents multimédia. Outil très général de représentation des connaissances doté d'une syntaxe XML, les cartes topiques sont articulées autour de quatre axes et ont pour but est d'agréger toutes les informations disponibles sur un sujet donné, et ce en articulant ces informations autour d'un point unique d'indexation (topic), en reliant ces points par un réseau sémantique de relations appelées associations.

A ce stade de notre étude, nous avons donc décrit les bases de la pyramide associée à la stratégie du web sémantique de Berner Lee.

Néanmoins, l'utilisation des descriptions de données par les langages RDF et RDFS ne peut à elle seule permettre au web sémantique d'exister. Nous l'avons vu, ces langages d'annotation permettent de mettre en forme les métadonnées des documents, mais ils se retrouvent vite confrontés à l'utilisation qui en est faite et particulièrement à la mise en place de réseaux sémantiques des métadonnées sur le web. Même si les métadonnées sont l'un des deux composants principaux du web sémantique, il convient donc d'explicitier quel est le second composant principal, et à quoi il sert : l'ontologie.

Les ontologies : formalisation et réseaux sémantiques

1. Définition et cycle de vie d'une ontologie

Le terme ontologie est initialement dédié à la philosophie et signifie *explication systématique de l'existence*, c'est-à-dire la doctrine ou théorie de l'être. Quel est alors le lien qui relie cette notion ancienne au web sémantique ? Concrètement, il n'y en a pas, mais de nombreux théoriciens, scientifiques ou savants se sont accaparés ce terme pour l'associer au domaine sur lequel ils travaillent. Que ce soit dans le domaine de la biologie, de l'informatique ou encore de l'intelligence artificielle, l'ontologie est alors considérée comme une forme de représentation de la connaissance au sujet du domaine (celui de la biologie, de l'informatique ou de l'intelligence artificielle) ou d'une certaine partie de ce monde. Ainsi, les ontologies décrivent généralement des individus (considérés comme les objets de base du domaine étudié), les classes (ensembles ou types d'objets), les attributs (propriétés que les objets possèdent et peuvent partager avec d'autres objets), les relations (liens entre objets) et enfin les événements (changements subis par des attributs ou des relations entre objets).

En informatique, la communauté définit l'ontologie comme un ensemble structuré des termes et des concepts représentant le sens (la sémantique...) d'un champ d'informations. Ce champs d'informations peut être par exemple des métadonnées ou d'autres éléments du domaine de connaissance choisi, en l'occurrence le web sémantique. L'ontologie est donc

une sorte de modèle de données qui représente un ensemble de concepts sur un domaine précis, mais également les relations associées à ces concepts.

La structure d'une ontologie, détaillée et élargie, est assez similaire à celle d'un dictionnaire. Elle permet aux machines de traiter son contenu de façon intuitive et automatique. Son apport, par rapport aux langages d'annotation de type RDF, réside dans le fait que l'ontologie va permettre de jouer réellement le rôle intelligent entre les données, leur sémantique, et leur compréhension par des agents logiciels ou des machines. Ainsi, par le biais d'une organisation en réseau sémantique, reliant, notamment grâce à RDF, les sémantiques de documents divers entre elles, l'ontologie va être l'outil indispensable à la création de valeur du web sémantique. Au-delà de la simple définition des concepts (principes, idées, catégorie d'objet...) et des relations entre ces concepts, l'ontologie va également organiser de façon hiérarchique ceux-ci et leurs idées, ainsi que les relations qui existent entre eux et les règles qui les contraignent. On comprends alors facilement pourquoi l'ontologie est importante dans le domaine de l'informatique et plus précisément du web sémantique.

Les ontologies représentent une des deux composantes majeures du web sémantique car celui-ci cherche justement à s'appuyer sur des modélisations des ressources du web afin de rendre celui-ci accessible par tous. Elles apportent une solution à cela en s'appuyant sur des représentations conceptuelles du domaine étudié, ce qui en fait un élément majeur dans la démarche proposée par Berner Lee.

L'ontologie en tant que résultat d'une formulation exhaustive et rigoureuse de la conceptualisation d'un domaine, apporte alors une solution idéale pour connecter entre eux les réseaux sémantiques associés à des contenus web.

En considérant l'ontologie comme un être vivant que l'on peut compléter et transformer de façon infinie, on peut alors faire émerger chaque étape de son cycle de vie : chacune d'elle pose des problèmes de recherche, il convient donc d'expliquer à quoi elles correspondent. Ce cycle de vie rassemble sept activités : détection des besoins, conception, gestion et planification, évolution, diffusion, utilisation, évaluation.

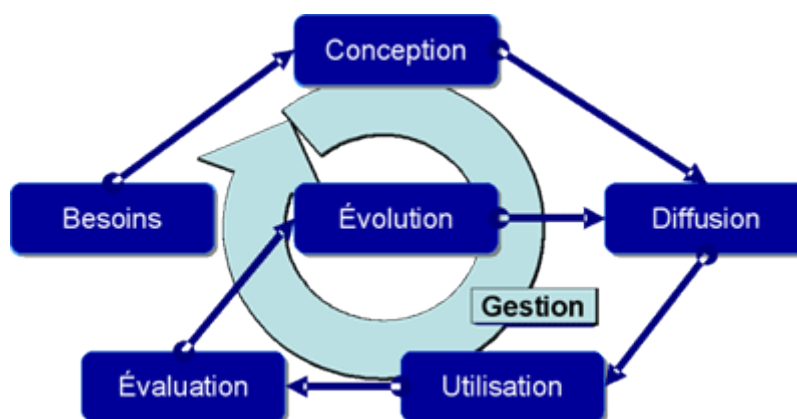


figure 5. Le cycle de vie d'une ontologie

- Besoins et évaluation

L'activité de détection des besoins précède la phase de conception et nécessite la mise en place de démarches méthodologiques de recueil d'information et d'identification de scénario potentiels. Cette phase de détection des besoins nécessite un état des lieux initial approfondi dans la mesure où elle ne peut reposer sur des études précédentes ou des retours d'utilisation, comme c'est le cas pour l'évaluation.

- Conception et évolution

La phase de conception de l'ontologie met en exergue un certain nombre d'actions à faire :

- Spécifier des solutions (maquettage, prototypage)
- Acquérir des connaissances nécessaires (traitement automatique de la langue naturelle, analyses de texte)
- Concevoir et modéliser (*design pattern* ontologiques, méta-ontologies)
- Formaliser l'ontologie (méthodes et outils de l'ontologie formelle, analyse formelle de concepts, graphes conceptuels, formalismes du web sémantique RDF/S et OWL) ;
- Intégrer des ressources existantes (alignement automatique d'ontologies, traduction) ;
- implantation (graphes conceptuels, logiques de description, formalismes objets).

Dans cette phase de conception, il s'agit également d'obtenir un consensus sur les choix de représentation et de conceptualisation de l'ontologie, car en fonction des usages, cette action nécessite des « collecticiels » et des outils de gestion de points de vue, de terminologies et des différentes langues.

Notons que la phase d'évolution de l'ontologie est liée à celle-ci car elle précède sa phase de diffusion : l'évolution pose le problème de la maintenance des contenus propres à l'ontologie existante. En effet, une ontologie est d'une part un objet vivant et également un ensemble de primitives facilitant la description des faits du domaine étudié. Lorsque l'ontologie est modifiée, les changements ont un impact sur ce qui avait été préalablement construit dans l'ontologie. La maintenance de l'ontologie soulève donc des dilemmes associés à l'intégration technique et l'intégration des usages.

- Diffusion de l'ontologie

Cette phase de diffusion de l'ontologie concerne le déploiement et la mise en place de celle-ci après l'avoir créée. Elle est souvent confrontée à des problèmes de compatibilité avec les architectures des solutions utilisées. Les architectures distribuées peuvent ainsi être utilisées pour le partage de fichier, celles de services web pour l'intégration d'applications... Ces différents types d'application nécessite une harmonisation dans la distribution des ressources (données, modèles...) et leur hétérogénéité au niveau syntaxique voir sémantique, implique une recherche sur l'interopérabilité de l'ontologie diffusée.

- Utilisation de l'ontologie

Cette phase du cycle de vie regroupe les activités relatives à la disponibilité de l'ontologie, par exemple, l'annotation des ressources par le biais de métadonnées, la résolution de requêtes sur cette ontologie, ou encore la déduction de connaissances et l'aide à la décision (moteurs d'inférence à base de règles). Ces types d'activités ont en commun la question de la conception des interactions avec l'utilisateur, ainsi que leur ergonomie.

- Gestion

Cette phase de gestion de l'ontologie est importante car elle met en valeur l'important travail de suivi pour détecter, déclencher, préparer et évaluer les itérations du cycle de

l'ontologie, en s'assurant que celle-ci reste située dans le cercle vertueux des systèmes d'information (contribution, utilisation, création)

2. Les Ontologies et le web sémantique

Les ontologies sont la technologie dorsale pour le web sémantique et pour le management des connaissances formalisées décrivant les ressources du web. La principale raison à cela est qu'elle fournissent la sémantique des données d'un document web, ainsi que les sources d'informations qui peuvent être transmises entre différents agents logiciels. Les métadonnées et annotations sémantiques décrivent les ressources en utilisant la sémantique définie dans l'ontologie. De ce fait, les ressources annotées via les métadonnées permettront de faciliter la recherche, l'extraction, l'interprétation et le traitement des informations du web de façon efficace. Le web sémantique est l'endroit idéal pour l'expansion des systèmes à base d'ontologies, car l'information va être associée en son sein à un sens bien défini : les agents logiciels vont alors pouvoir traiter les informations des contenus web de façon optimisée. Les descriptions de données offertes par les ontologies permettent d'améliorer les fonctionnalités des agents logiciels.

Grâce au web sémantique, l'ontologie a donc trouvé un certain formalisme à l'échelle mondiale et s'intègre de façon logarithmique dans les applications web, la plupart du temps de façon transparente pour l'utilisateur. Les secteurs où les ontologies de domaines sont disponibles ne cessent donc de s'étendre : ontologie de la génétique, ontologie de la géométrie, ontologie pour les musées, ontologie médicale, ontologie pour l'enseignement, ontologie pour le bâtiment, ontologie de systèmes documentaires, ontologie pour la gestion, ontologie dans le secteur automobile... et cette expansion ne semble pas être finie à l'heure où l'ontologie se dote d'une ingénierie, preuve en est l'ontologie qui s'appliquait essentiellement à des données (images, documents...) est dorénavant utilisée pour décrire des logiciels ou services web, ainsi que leur caractéristiques fonctionnelles et non fonctionnelles.

De même, l'ontologie commençait déjà à être utilisée pour décrire les utilisateurs et s'étend maintenant à la description du contexte d'interaction, pour doter les applications de ce que l'on appelle une « conscience du contexte ». Cela concerne les préférences de l'utilisateur (langue, goûts, droits, etc.), les caractéristiques du terminal (mobile, vocal, etc.), la situation géographique (à l'étranger, dans une salle avec imprimante, etc.), l'activité en cours (au volant, en présentation, etc.), l'historique d'utilisation. Enfin, si l'ontologie est actuellement utilisée pour faciliter l'accès à des informations et des applications, on pressent aussi son utilisation dans la description et l'application de règles de sécurité et de confidentialité décrites à de hauts niveaux d'abstraction, permettant de restreindre les accès avec une grande flexibilité. Ainsi, dans un système d'information à base d'ontologies, la confidentialité et les règles reposent aussi sur la sémantique d'une ontologie et les inférences qu'elle permet pour contrôler l'accès à l'information et la précision de l'information diffusée.

3. Les langages de définition d'ontologies

Pour exister, les ontologies ont besoin d'être définies par des langages standardisés spécifiques : le langage principal est OWL (Web Ontology Language) et permet de définir et

représenter des ontologies , il est dérivable en 3 sous langages : OWL LITE, OWL DL et OWL FULL.

La syntaxe d'un document OWL est définie par celle des différents constructeurs utilisés dans ce document. Elle est d'ailleurs le plus souvent donnée sous la forme de triplets RDF. On pourra y voir une association entre chaque constructeur et la sémantique issue des logiques de descriptions. En outre, la sémantique rattachée au mots clés du langage OWL est plus précise que celle associée au document RDF : la mise en place de l'ontologie permet de rendre la sémantique des contenus web accessible via des concepts intelligent, l'ontologie va par exemple permettre de faire des déductions à partir de la sémantique originale d'un document.

En effet, le langage RDF, focalisé sur les métadonnées des relations entre objets, s'est trouvé limité quant à la nécessité d'élaborer une définition des propriétés et des classes dont les objets sont des instances. L'extension RDF Schéma fournit certes des mécanismes basiques pour spécifier ces classes, mais cela ne s'avère pas suffisant pour permettre d'implémenter correctement la sémantique associée au contenu web.

Le langage OWL, tout comme RDF, est un langage profitant de l'universalité syntaxique du XML. Il est spécifiquement dédié aux définitions de classes et de types de propriétés, il est donc directement associé à la définition d'une ontologie. Il utilise un vocabulaire et une sémantique formelle qui lui permettent d'offrir aux machines des capacités supérieures pour interpréter les contenus web. La particularité d'OWL est qu'il s'inspire des logiques de descriptions en fournissant un grand nombre de constructeurs, qui eux-mêmes permettent d'exprimer, d'une part les classes correspondant aux connecteurs de la logique de description équivalente, d'autre part les propriétés des classes définies, leur cardinalité et des caractéristiques des propriétés. Comme précisé précédemment, le langage OWL se compose de trois sous-langages qui symbolisent une expressivité croissante.

- OWL-Lite ne contient qu'un sous-ensemble réduit des constructeurs disponibles. Il est le sous langage de OWL le plus simple, destiné à représenter des hiérarchies de concepts simples. Il a la complexité formelle la plus basse et l'expressivité minimale dans la famille OWL.

- OWL-DL contient l'ensemble des constructeurs, mais avec des contraintes particulières sur leur utilisation qui assurent la décidabilité de la comparaison de types. Fondé sur la logique descriptive d'ou son nom (OWL Description Logics), il est adapté pour faire des raisonnements, et il garantit la complétude des raisonnements et leurs décidabilité.

- OWL-Full est conçu pour ceux qui ont besoin de l'expressivité maximale, de la liberté syntaxique de RDF mais sans garantie de calculabilité. Destiné aux situations où il est important d'avoir un haut niveau de capacité de description, quitte à ne pas pouvoir garantir la complétude et la décidabilité des calculs liés à l'ontologie

Les ontologies OWL sont présentées sous la forme d'un fichier texte dont l'extension peut être .owl ou .rdf. Ceci est notamment dû au fait qu'OWL repose sur RDF et RDFS, en y ajoutant notamment des nouvelles balises pour gagner en précision.

Un langage de requête indispensable

Une ontologie OWL est comparable à une base de données relationnelle, dans la mesure où elle contient toute la sémantique du document en suivant une formalisation bien définie. Légitimement, on pourrait se demander quel est le langage qui va permettre de faire des requêtes sur cette ontologie, à l'instar du langage SQL pour une base de données, ou de Xpath pour les documents structurés.

La capacité de réaliser des requêtes sur les connaissances représentées au sein du web sémantique joue un rôle primordial dans l'accès aux applications et bases de connaissances. Le langage de requête SPARQL a été proposé par le W3C pour être le langage dédié à RDF. Il est en outre très utilisé dans le domaine de la recherche d'extractions d'informations. Il est principalement basé sur la correspondance des patrons de graphes (matching graph patterns) comme le patron de triplet RDF. D'autre part, SPARQL intègre également des balises (tags) spécifiques telles que le patron de graphe optionnel, l'union et l'intersection des patrons, le filtrage, les opérateurs de comparaison des valeurs permettant d'effectuer des requêtes plus efficaces et flexibles.

```
1. PREFIX dc: <http://purl.org/dc/elements/1.1/>
2. PREFIX ns: <http://example.org/ns#>
3. SELECT ?name ?price
4. WHERE { ?x ns:name ?price .
5. ?x dc:name ?price .
   FILTER (?price < 30) .
6. }
```

Cet exemple de requête en SPARQL demande le nom et le prix des ressources dont le prix est inférieur à 30 €.

Langages de description et de composition de service

Certains organismes industriels ou universitaires proposent régulièrement des langages dédiés aux services web. Il convient ici, dans le cadre de l'état de l'art des technologies du web sémantique, de les mentionner rapidement, sachant que ces langages peuvent être complémentaires et ne répondent pas nécessairement aux mêmes besoins

1. UDDI

Le protocole UDDI (Universal Description, Discovery and Integration) est une plateforme dont le rôle est de stocker les descriptions de services web disponibles à la manière d'un annuaire.

2. E-Speak

Initialement développé par Hewlett-Packard, ce langage fût un des pionniers en matière de service web dès 1995. Les buts d'E-speak étaient sensiblement les mêmes que ceux d'UDDI (publication de services), mais avec une approche plus générale.

3. WSDL

Le langage WSDL est basé sur XML et permet de décrire des interfaces de services web. Il permet la représentation abstraite des opérations que les services peuvent réaliser, et ce de façon indépendante de l'implémentation qui en a été faite. Langage ne comportant pas de moyen de décrire de manière plus abstraite les services (comme UDDI), ni de moyen de conversation et de transaction de messages (tel SOAP), il est néanmoins utilisé de manière générale en tant que passerelle entre ces représentations de haut niveau et de bas niveau.

4. DAML-S

Ce langage permet de décrire des services en utilisant le modèle logique de description de ses services. Basé sur XML, son principal intérêt est qu'il est un langage de haut niveau pour la description et l'invocation des services web dans lequel la sémantique est comprise.

5. OWL-S

OWL-S (Web Ontology Language for Services) sert à identifier des web services. C'est un schéma OWL basé sur une ontologie. Il est le descendant de DAML-S.

6. XL

Plateforme axée sur XML, elle est destinée au services web et utilise son langage propre (XL) tout en prenant en compte les technologies du W3C de type WSDL et SOAP afin de rendre possible l'interopérabilité de ses applications avec d'autres applications.

7. ebXML

Ce standard (Electronic Business XML initiative) développé en 1999 est orienté vers le commerce électronique et est très industriel. Il facilite la standardisation des échanges pour le commerce inter-entreprise, en remplacement du standard actuel EDI.

8. RosettaNet

RosettaNet est assez similaire à ebXML. Néanmoins, ebXML correspond à une conception horizontale des services web, alors que RosettaNet en a une vision verticale : les informations échangées le seront (entre plusieurs entreprises ou au sein d'une même entreprise) en général concernant un produit ou un ensemble de produits bien particuliers. RosettaNet est actuellement utilisé par plus de 400 acteurs majeurs de l'industrie de l'électronique et des semi-conducteurs, des SSII et des entreprises de Technologies de l'Information.

Réflexion analytique

1. Comparaison du web 2.0 au web sémantique

Tout d'abord, le web 2.0 et le web sémantique sont-ils comparables ? De nombreuses rumeurs et idées circulent sur le sujet.

Suite à cette recherche, nous allons vous présenter différents avis ou différentes analyses.

Nous proposons tout d'abord une comparaison des deux webs, proposée sur le site Acidweb⁴. Cet article reprend l'explication d'économistes dans le site O'reilly radar⁵ de Tim O'Reilly, fondateur et CEO d'O'Reilly Media, Inc.

« Pour le web sémantique, la méthode consiste à rendre internet intelligible par les moteurs de recherche. Pour cela on ajoute des balises de code qui renseignent les moteurs sur les relations sémantiques et le sens créé par leur combinaison. La technologie déployée serait alors capable de comprendre le sens de ce qui est écrit, alors qu'aujourd'hui elle ne fait qu'en répertorier les mots. Ainsi du point de vue sémantique du web, la toile est dénuée de signification, et il reste à créer les moyens de la comprendre.

Le web2.0 a une approche opposée: le sens de l'information est déjà présent, de manière intrinsèque, mais il reste à exploiter. L'information n'a de valeur que dans l'usage qui en est fait, et c'est à travers l'utilisation par des millions d'internautes qu'elle va acquérir son sens. Le web2.0 est une approche pragmatique de la caractérisation du web: la signification est là, brut de décoffrage, et c'est l'usage des utilisateurs qui va la sculpter et lui conférer sa valeur.

Pour illustrer ce concept, Tim revient sur la naissance du page rank. Les webmasters ont une tendance naturelle à poser des liens sur leur site. Ces liens pointent vers d'autres sites qui sont généralement en rapport avec le contenu du premier. La grande intelligence de Larry Page a été de considérer ces liens comme des votes de pertinence pour perfectionner l'algorithme Google. Si un site est lié par un grand nombre de sites qui abordent des thèmes connexes, cela signifie qu'il est reconnu par ses pairs, et qu'il y a de grandes chances pour que l'information traitée soit très pertinente. L'algorithme, à travers le « page rank » prend cette reconnaissance en compte et « booste » son référencement naturel. Larry Page n'a pas inventé le « linkin »g entre les sites. Il a simplement accepté de le comprendre différemment, de lui donner un nouveau sens.

Le phénomène est le même pour le web2.0: le sens de l'information est présent: il suffit de le faire émerger par l'utilisation que l'on fait, collectivement, de cette information. »

Nous ne pouvons pas parler de comparaison du web 2.0 et du web sémantique sans citer Luciano Floridi, chercheur en philosophie de l'information de l'université de Hertfordshire. En effet Luciano Floridi a écrit un article intitulé : « Web 2.0 vs. the Semantic Web: A Philosophical Assessment »⁶, traduit en français par Patrick Peccatte.

⁴ <http://www.acidweb.fr/web20/web20-vs-web-semantique>

⁵ <http://radar.oreilly.com/archives/2007/09/economist-confu.html>

⁶ <http://www.philosophyofinformation.net/publications/pdf/w2vsw.pdf>

Dans cet article, il examine le développement du web sémantique et du web 2.0 dans le but d'en déterminer leur avenir. Il soutient l'idée que le web sémantique est un projet clair et bien défini mais que malgré cela il ne constitue pas une réalité prometteuse et qu'il échouera de la même façon qu'a échoué le projet de l'Intelligence Artificielle dans le passé. Il soutient aussi que le web 2.0, bien qu'il soit assez mal défini et qu'il manque une explication claire de sa nature et de sa portée, a la capacité d'être un véritable succès.

Il explique différentes raisons qui font que le web sémantique « *ne marchera pas* ». Tout d'abord selon lui, le web sémantique est confus sur un plan conceptuel. Les concepts clés le décrivant ("sémantique", "sens", "compréhension", "information", "connaissance" et "intelligence",) «*sont tous utilisés à mauvais escient, employés trop vaguement ou tout simplement métaphoriques.* ». Les langages, protocoles et ontologies pour les métadonnées et la métasyntaxe permettent « *l'intégration, l'agrégation, le partage, la syndication et l'interrogation de données hétérogènes, mais néanmoins bien circonscrites à un sujet et à travers différentes bases de données* ». Pour lui il n'est pas question ici de sémantique.

Dans un autre temps, il présente aussi le web sémantique comme une version forte du projet d'intelligence artificielle, ce qui pour lui signifie qu'il est irréalisable. Pour résumer ses propos, il fini par : « *Nous devrions donc en réalité parler de Web lisible par une machine ou de Web des données comme le fait le W3C.* »

Malgré les lectures que nous avons eu sur le sujet, nous ne trouvons pas pertinent de prendre parti et de comparer les deux web, car de notre point de vue, il ne sont pas comparables de part leurs structures et les idéologies qu'ils ont développé.

Nous allons cependant en faire une description sémantique et linguistique en présentant les types de recherches des deux web : la recherche par mots clés du web 2.0 et la recherche sémantique du web sémantique.

2. Moteurs de recherche

Les moteurs de recherche actuels

Les moteurs de recherche actuels sont des moteurs de recherche lexicaux. Ils permettent à un internaute de trouver les informations qu'il cherche sur le web à partir d'une requête. Le moteur est constitué d'une base de donnée appelée « Index » où sont stockées les informations présentes sur Internet. Le moteur compare la requête formulée par l'utilisateur aux documents conservés dans son index. Lorsqu'il y a des correspondances, la page est sélectionnée puis présentée à l'internaute. Cependant, le moteur de recherche procède d'abord au « Ranking » des pages pour les classer par pertinence grâce à des algorithmes tenus secret.

On peut noter trois phases importantes effectuées par le moteur de recherche : le crawl (le moteur de recherche consulte des pages et collecte des données), l'indexation (les données sont enregistrées dans l'index du moteur) et le ranking (notation des pages).

Les moteurs de recherche actuels sont donc basés sur le principe de la recherche par mots clés, appelée aussi recherche lexicale. La taille de l'index du site est proportionnelle à la pertinence des résultats obtenus, mais à partir d'une certaine taille, l'index est pollué par

trop d'informations. Ce qui entraîne un manque de pertinence, et un déficit de productivité côté utilisateur.

Pour ce type de moteur de recherche l'indexation se fait donc sur l'ensemble du texte, en tenant compte de la location et des occurrences des mots recherchés. Il existe en suite des techniques de tri pour améliorer la pertinence des résultats obtenus.

Pour la notation et le classement des pages, le moteur de recherche le plus utilisé, Google, a fortement utilisé (et utilise toujours mais beaucoup moins) le Page Rank (PR). C'est un procédé qui permet de fixer l'importance d'un site en fonction de sa popularité sur le web. Chaque page web sur lequel se trouve un hyperlien pointant vers un site est un vote pour ce site. Donc plus un site possède est référencé sur internet, plus son Page Rank est élevé. Ce résultat du Page Rank est pris en compte dans le tri des pages web à afficher lors d'une recherche. Donc si cet indicateur est grand, le site sera plus visible dans les résultats de Google. Les sites n'ont pas le tous le même poids : si un site a un RP élevé, son lien vers un autre site aura plus de poids qu'un deuxième site dont le RP est faible .



Figure 6 : Illustration du Page Rank




Le clustering

Dans le domaine des moteurs de recherche, le terme clustering regroupe deux notions. La première notion correspond au fait que le moteur de recherche ne peut donner que deux/trois pages différentes d'un même site.

[Semoule - Toutes les recettes de cuisine - auFeminin.com](#) ★

Semoule - Toutes les recettes de cuisine. ... Recettes à base de **Semoule** les plus appréciées : ... Nouilles orientales à la **semoule** et au safran, 30 mn ...




www.aufeminin.com/.../semoule.html - Il y a 8 heures -

[En cache](#) - [Pages similaires](#) -   

[Recette : Semoule au lait - Recette auFeminin](#) ★

2-Verser la **semoule** en pluie dans le lait bouillant, mélanger. 3-Ajouter le beurre coupé en petits morceaux et mélanger encore. ...

www.aufeminin.com/.../semoule-au-lait.html - Il y a 7 heures -

[En cache](#) - [Pages similaires](#) -   

Inclassables : [Semoule ultra rapide pour le couscous - Recette de ...](#) ★

Figure 7 : Exemple de Clustering sur Google

La première page apparaît normalement tandis que la seconde est en dessous et décalée vers la droite. Tous les moteurs utilisent ce type de clustering pour éviter qu'un site ne prenne toutes les places sur une pages de résultats.

La seconde signification désigne le fait de ranger les résultats en dossiers thématiques. Basée sur l'apprentissage non supervisé, cette classification regroupe les résultats sur une base de similarités trouvées dans tout ou seulement une partie des documents. Via un algorithme, les différents groupes sont créés et les résultats sont rangés par pertinence au sein même du groupe. La relation de ressemblance et de similitude entre les documents est fondée sur des caractéristiques statistiques et sémantiques.

Cette méthode n'est pas utilisée par tous les moteurs de recherche mais elle est applicable sur les moteurs de recherche par indexation et les moteurs de recherche sémantiques.

Moteur de recherche encyclopédique : Wolfram Alpha

Stephen Wolfram, chercheur britannique, a lancé en mai 2009 le moteur de recherche en langage naturel Wolfram Alpha. Ce spécialiste de la modélisation a déjà développé le logiciel Mathematica, sur lequel s'appuie son moteur de recherche. Lors d'une requête, Wolfram Alpha n'apporte pas de liens comme le font les moteurs de recherche actuels, mais directement une réponse précise à la question posée à partir d'une base de connaissance. Pour le moment il ne prend en compte que les requêtes rédigées en anglais et ne couvre pas toutes les thématiques. Stephen Wolfram a promis de compléter sa base de connaissance dans les mois et les années à venir.

Les technologies utilisées pour ce moteur de recherche ne sont pas celles du web sémantique : « *La technologie employée par Wolfram Alpha diffèrerait cependant des standards sémantiques tels que l'OWL (Web Ontology Language) ou le RDF, et aurait recours à des millions de lignes d'algorithmes...* »⁷

Nous proposons un exemple d'utilisation de Wolfram Alpha. Nous interrogeons le moteur avec la requête suivante : « Barack Obama birth day ».

⁷ <http://eranksolution.net/blog/2009/05/wolfram-alpha-moteur-en-langage-naturel/>

WolframAlpha™ computational knowledge engine

Barack Obama birth day

Interpreting "Obama" as "obomba"

Input interpretation:

Barack Obama	date of birth
Barack Obama	

Results:

date of birth	August 4, 1961 (age: 48 years)
date of birth	August 4, 1961 (age: 48 years)

Computed by: [Wolfram Mathematica](#) [Source information »](#) Download as: [PDF](#) | [Live Mathematica](#)

New to Wolfram|Alpha?

A few things to try:

- enter any date (e.g. a birth date)
june 23, 1988
- enter any city (e.g. a home town)
new york
- enter any two stocks
IBM Apple
- enter any calculation
\$250 + 15%
- enter any math formula
 $x^2 \sin(x)$

[more »](#)

[Examples by Topic »](#)
[Visual Gallery of Examples »](#)
[Watch Overview Video »](#)

La réponse proposée est sous forme d'un table nous donnant la date de naissance exacte de Barack Obama ainsi que son âge.

Nous avons décidé d'effectuer quelques tests de cet outil pour des recherches spécifiques. Si nous cherchons le terme « ANOVA »⁸ dans ce moteur de recherche, la réponse obtenue est que l'anova est un gène et que sa fonction est « RNA binding ». Wolfram Alpha détaille alors quelques notions de neuro-oncologie. Cette même recherche dans google nous permet de savoir que l'anova est un test statistique et nous renseigne sur les méthodes théoriques de l'anova. Nous avons ensuite utilisé google pour savoir si l'anova correspond aussi à un nom de gène. Lors de cette recherche, nous avons trouvé sur les 5 premières pages des textes sur la façon d'utiliser l'anova pour lister, sélectionner ou comparer les gènes. Ce n'est qu'à la page 5 que nous obtenons un site sur la neurologie et le gène anova. Pour approfondir la recherche sous google, nous introduisons le mot RNA dans notre requête. Les liens obtenus traitent bien de neurologie et du gène anova.

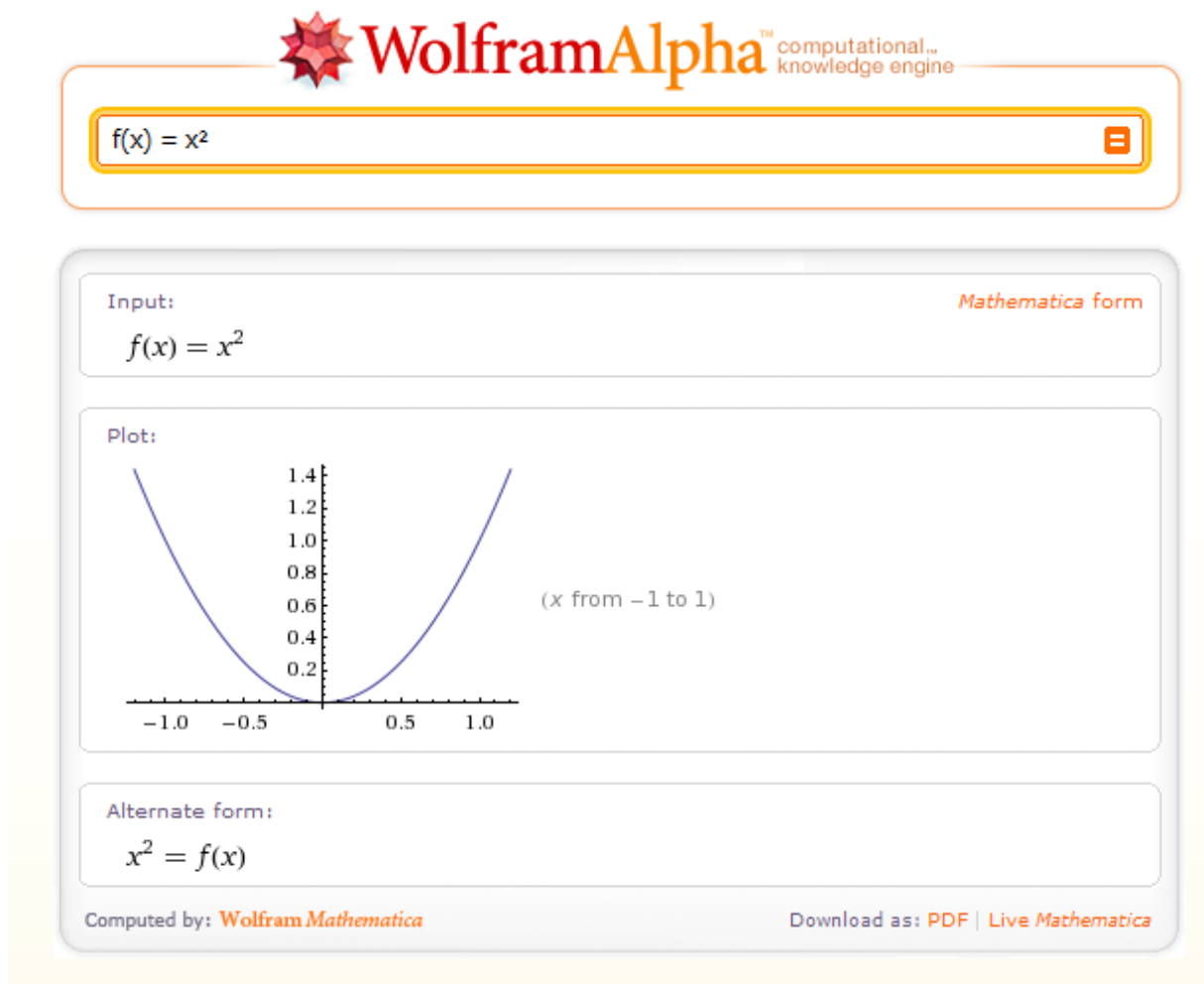
Bien que très utilisé dans le monde de la statistique, le terme anova est une abréviation. Nous avons donc tenté une nouvelle recherche avec le terme complet : « Analysis of variance ». Sur google, nous obtenons encore des liens vers des sites expliquant l'utilisation et l'application de l'analyse de la variance. Wolfram alpha reconnaît bien l'analyse de la variance et nous donne une définition de cette analyse. Il nous propose un synonyme : l'Anova et nous indique que c'est un terme proche de l'analyse multivariée. Aucune méthodologie de ce test, ni explication plus approfondie.

On peut donc s'interroger sur les différentes réponses obtenues via Wolfram Alpha et google. Est-ce pertinent de retourner des informations sur le gène anova lors d'une requête sur le terme anova ? Le pourcentage de personne cherchant des informations sur ce gène à partir de ce terme est tout de même très faible. La plupart des personnes cherchent des données sur l'analyse de la variance.

⁸ Abréviation d'ANalysis Of VAriance

De plus, Wolfram Alpha ne retourne que des informations succinctes sur le thème demandé. Ceci nous mène à conclure que Wolfram Alpha n'est pas comparable aux moteurs de recherche. Il ressemble plus à une encyclopédie. Cet outil permet de répondre à des questions concrètes, basées sur des faits.

Cependant il a une toute autre fonctionnalité que ne proposent pas les différents moteurs que l'on trouve actuellement. Il peut effectuer de nombreux calculs et nous donner des analyses statistiques. Par exemple, si je fais une recherche avec pour requête « $f(x)=x^2$ », voici ce que me propose Wolfram Alpha :



Ou encore, il peut résoudre des équations :

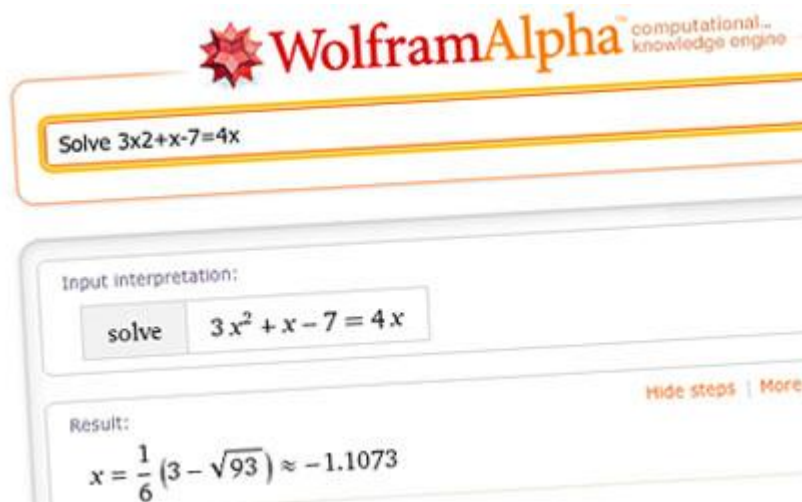


Figure 8 : équation

<http://www.tunibox.com/divers/wolfram-alpha-le-premier-moteur-semantique-qui-revolutionne-le-web.html>

On pourrait encore montrer de nombreuses applications de Wolfram Alpha tant il y en a. Pour conclure sur les capacités de cet outil, on peut dire qu'une fois qu'on en a déterminé les applications possibles, il paraît puissant. Ce n'est clairement pas un moteur de recherche au sens google, ni une encyclopédie comme wikipédia. C'est un mélange savant des deux. Un outil qui permet de requêter de manière intelligente une base de données encyclopédique. Dans les domaines scientifiques ou les domaines demandant des données factuelles et précises (tels que la géographie, l'histoire, la démographie, etc.), il est facile d'obtenir une réponse claire et juste. Wolfram Alpha est un moteur de recherche à prendre en complément de google (ou autre moteur de recherche).

Que peut-on dire de son positionnement par rapport au web sémantique ? Même si certaines finalités sont les mêmes, Wolfram Alpha n'utilise pas les techniques du web sémantique. Pourtant, nombreux sont les articles que l'on trouve sur internet, présentant Wolfram Alpha comme un nouveau moteur de recherche sémantique et une révolution pour le web sémantique. Nous pensons que cet outil est encore trop jeune pour pouvoir en faire une analyse et le situer par rapport au web sémantique. En effet, il est difficile de trouver de la documentation pertinente sur ce sujet.

Ce moteur de recherche intrigue et intéresse. Il serait question d'une intégration de Wolfram Alpha dans le moteur de recherche Bing de Microsoft.

Les moteurs de recherche sémantiques

Les moteurs de recherche sémantiques sont basés sur la reconnaissance des concepts, l'identification du sens des mots : le Traitement Automatique du Langage Naturel (TALN). Celui-ci a pour objet le développement de programmes informatiques dont le but est de traiter automatiquement les langues naturelles. Une langue naturelle désigne une langue parlée ou écrite par des êtres humains, contrairement aux langages artificielles (par exemple, langage informatique, mathématiques ou logistique). Parmi les applications les plus connues, on peut citer : la traduction automatique (historiquement la première application, dès les années 1950), la correction orthographique, la recherche d'information et la fouille de textes, le résumé automatique de texte, la génération automatique de textes, la synthèse de la parole, la reconnaissance vocale, la détection de registre, la

classification et la catégorisation de documents, la reconnaissance de l'écriture manuscrite, les agents conversationnels.

Le TALN s'appuie sur les principes suivant :

- L'analyse morphologique
- L'analyse lexicale
- L'analyse syntaxique
- L'analyse sémantique

Pour comprendre l'action d'un moteur de recherche sémantique, il est nécessaire d'expliquer quelques notions de sémantique linguistique.

La sémantique est l'étude du sens des unités linguistiques et de leurs combinaisons. Le champ sémantique est « *l'ensemble des différentes significations d'un même mot dans les différents contextes où il se trouve* »⁹. Le champ lexical est quant à lui l'ensemble des mots d'un texte qui se rapportent à une même réalité ou à une même idée. Ces mots sont des synonymes, appartiennent à la même famille, au même domaine, ou encore renvoient à la même notion. Le champ lexical renseigne donc sur le thème du texte.

L'analyse morphologique consiste à découper les mots en éléments plus petit, appelés morphèmes, mettre en évidence ces signes minimaux et à analyser la façon dont ils peuvent constituer des mots. On peut définir trois types de connaissances morphologiques :

- La morphologie flexionnelle : qui est l'étude de la formation des mots composés d'affixes apportant une information supplémentaire mais qui ne changent pas la catégorie syntaxique du mot. Par exemple, le suffixe « s » de formation des mots en français (fille + s -> filles).
- La morphologie dérivationnelle : qui est l'étude de la formation des mots composé d'affixes qui peuvent changer la catégorie syntaxique du mot. Par exemple celui de la formation des adverbe : « ment » ou « ement ». (probable + ment -> probablement).
- La morphologie : qui est l'étude de la combinaison de différents mots pour créer des mots nouveaux.

Exemple d'analyse morphologique sur le mot « inimitables » :

⁹ <http://www.espacefrancais.com/notion/c.html>

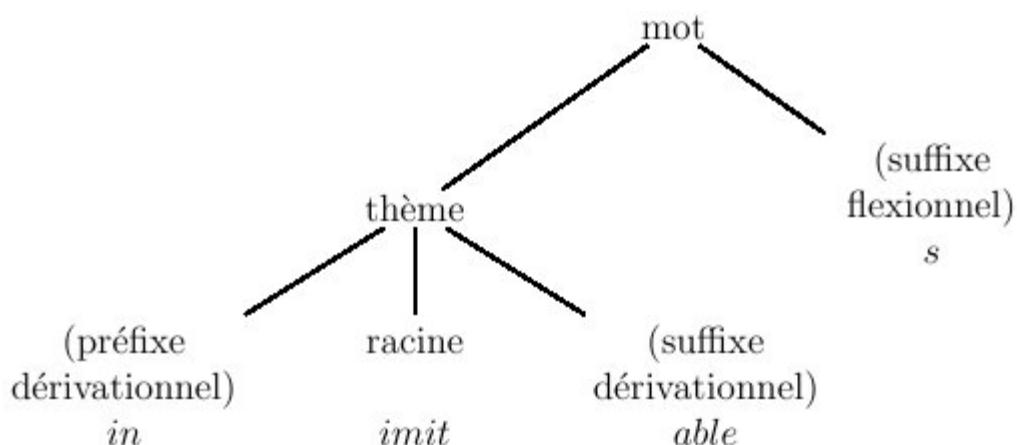


Figure 9 : Affixation

L'analyse lexicale est proche de l'analyse morphologique à la différence qu'elle s'intéresse au mot dans sa globalité. Elle permet de déterminer les mots proches du mot étudié (mot de la même famille, synonymes..) et de définir le lexique du mot. Grâce à l'analyse lexicale on peut déterminer le type de langage utilisé (familier, courant, soutenu,..)

Une fois les mots identifiés, on s'intéresse à la structure syntaxique des phrases. L'analyse syntaxique correspond à une analyse grammaticale. Elle décrit la manière dont sont combinés les mots formant ainsi des phrases syntaxiquement correctes. Cette analyse permet de définir la structure grammaticale d'un mot et son rôle au sein de la phrase.

Exemple d'analyse syntaxique sur la phrase « le chat mange une souris dans son panier ».

Le	Chat	Mange	Une	Souris	Dans	Son	Panier
Déterminant (Det)	Nom	Verbe (Vb)	Déterminant (Det)	Nom	Préposition (Prep)	Déterminant (Det)	Nom

On détermine ensuite le rôle des mots (ou groupe de mots) dans la phrase :

(Le chat)_{GN} ((mange (une souris)_{GN})_{GV} (dans (son panier)_{GN})_{GP})_{GV}.

Avec : GN= groupe nominal, GP = groupe prépositionnel et GV = groupe verbal.

Ce qui nous permet d'obtenir le graphe suivant :

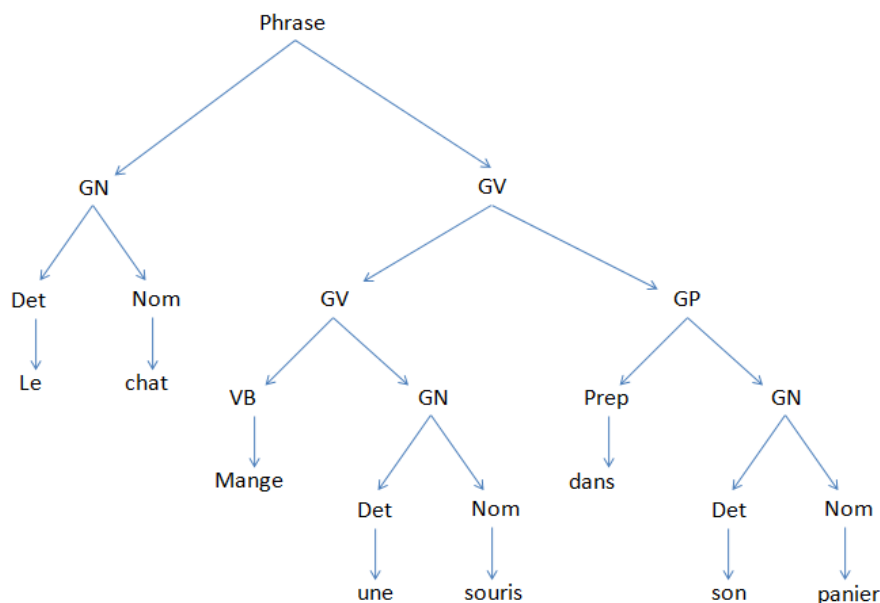


Figure 10 : Représentation syntaxique

Enfin, l'analyse sémantique permet de déterminer le sens des mots, et le sens globale d'une phrase. Elle permet d'étudier la manière de combiner les mots pour former une phrase dotée de sens.

La recherche sémantique prend bien sûr en compte

- La graphie, c'est-à-dire un correctif des fautes d'orthographe et de frappe via une recherche phonétique et un correcteur orthographique
- La polysémie via une analyse syntaxique et grammaticale
- Les expressions composées et les synonymes grâce aux ontologies

Connaître la qualité d'un moteur de recherche ?

Quatre critères permettent de caractériser la qualité d'un moteur de recherche: la précision, le rappel, le bruit et le silence. C'est le principe de la classification supervisée qui nous permet de calculer ces indicateurs.

Lorsque l'on fait une requête sur internet, les résultats obtenus ne sont pas toujours pertinents. Cette notion de pertinence est une notion imprécise qui dépend de l'interprétation de chacun. C'est cependant ce concept qui nous permet de calculer des indicateurs de qualité pour des logiciels tels que les moteurs de recherche.

On peut définir la pertinence comme étant :

- la correspondance entre un document et une requête
- le degré de relation (chevauchement, relativité, ...) entre le document et la requête

- le degré de surprise qu'apporte un document ayant un rapport avec la requête de l'utilisateur
- une mesure d'utilité du document pour l'utilisateur;

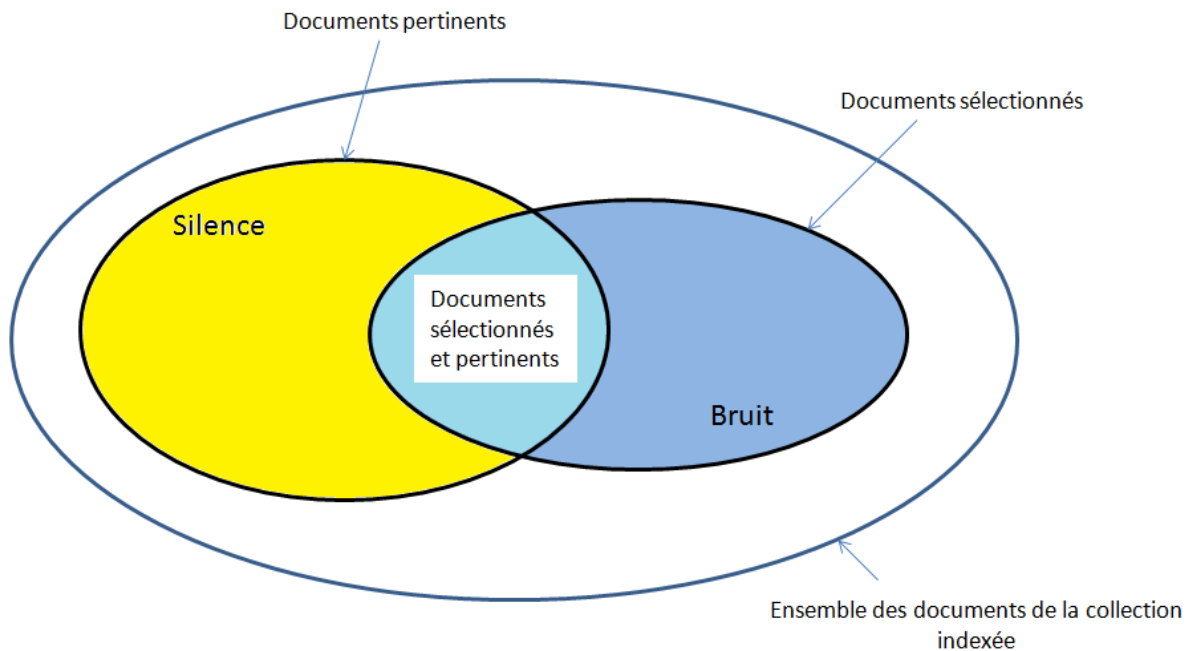


Figure 11 : Bruit et silence

Le bruit correspond aux documents retournés par le moteur de recherche mais non pertinents. Le silence correspond aux documents pertinents contenus dans l'index du moteur de recherche mais non retournés lors de la requête.

Le rappel est la proportion de documents retournés dans l'ensemble des documents pertinents.

$$\text{Rappel} = \frac{\text{retournés et pertinents}}{\text{pertinents}} = \frac{\text{corrects}}{\text{pertinents}}$$

La précision est la proportion de documents pertinents parmi les documents retournés.

$$\text{Précision} = \frac{\text{retournés et pertinents}}{\text{retournés}} = \frac{\text{corrects}}{\text{pertinents}}$$

Ce qui permet de calculer le silence qui est la proportion de documents manqués dans l'ensemble des documents pertinents.

$$\text{Silence} = 1 - \text{Rappel} = \frac{\text{Faux négatifs}}{\text{Pertinents}}$$

Le bruit est la proportion de documents non pertinents parmi les documents retournés.

$$\text{Bruit} = 1 - \text{Précision} = \frac{\text{Faux positifs}}{\text{retournés}}$$

3. Les applications du web sémantique

L'utilisation des technologies du web sémantique est à l'origine de nombreux types d'applications :

- les applications d'entreprises : capitalisation des connaissances, mutualisation des données, etc.
- l'e-formation ou e-learning: les technologies du web sémantique et les ontologies facilitent l'accès aux ressources pédagogiques disponibles sur le web. Dans le e-learning, ces technologies peuvent aider un enseignant à construire un cours, et à un étudiant à construire un curriculum personnalisé.
- l'e-commerce ou e-entreprise: personnalisation de la relation client avec un profil associé et services disponibles afin de guider l'achat ou le service après-vente, orchestration des processus métiers (*Business Process Management, Workflows*), etc.
- les loisirs : partage, communication, collaboration et échange pour le grand public.
- les sciences : plus particulièrement les sciences du vivant.

Nous allons détailler certaines de ces applications.

L'e-learning

Le e-learning, ou apprentissage en ligne, correspond à un mode de formation par des moyens électronique. Il a été défini en commission européenne comme « *l'utilisation des nouvelles technologies multimédias et de l'Internet, pour améliorer la qualité de l'apprentissage en facilitant l'accès à des ressources et des services, ainsi que les échanges et la collaboration à distance* ».

On trouve sur le web énormément de ressources utiles aux enseignants comme aux élèves : transparents, cours électroniques, bibliographie de documents auxquels les cours font référence, exercices, FAQs et toutes sortes d'autres documents pédagogiques. Dans cette quantité de ressources, il est difficile de s'y retrouver et d'accéder à des documents pertinents. Le web sémantique intervient une première fois à ce niveau pour le e-learning.

Les technologies du web sémantique permettent aussi de classer et d'indexer les documents afin de les mettre en ligne ou dans une bibliothèque privée. Les méta-données et les annotations sont donc utilisées pour classer les données sur le web. Les méta-données sont traitables par des machines, ce qui sert au e-learning. Le matériel du e-learning est sémantiquement annoté. La navigation à travers ce matériel pédagogique est donc possible avec les ontologies. Elles permettent aussi d'indexer les documents de façon à prendre en compte les caractéristiques des utilisateurs et de leur donner des documents qui répondent à leurs attentes et à leur profil.

Plusieurs aspects montrent que les ontologies peuvent s'appliquer au e-learning

- Les échanges de connaissances
- La conception et la maintenance de mémoires : indexation du matériel, requêtes sur le matériel, etc.
- Une connaissance du monde de l'enseignement

Les ontologies offrent :

- Une solution logicielle pour réduire la confusion conceptuelle et terminologique
- Une plus grande réutilisation
- Un partage plus large
- Une interopérabilité plus étendue

Il existe donc des standards pour décrire le matériel pédagogique tels que le standard IEEE LOM, IMS, etc.

Nous allons vous présenter quelques projets en cours de e-learning sémantique.

Trial solution

C'est un projet européen de Michel Buffa¹⁰, Sylvain Dehors¹¹, Catherine Faron-Zucker⁹ et Peter Sander⁹, qui permet la publication de documents personnalisés à partir de livres scientifiques. Les livres sont découpés en « ressources pédagogiques » et sont annotés avec des métadonnées pour décrire les relations entre les ressources ainsi que leur contenu. Une fois découpée et annotée, la ressource est disponible sur un serveur où les utilisateurs peuvent se servir. Ils utilisent ces données pour former par la suite leur propre document.

QBLS

QBLS, créé par Sylvain Dehors¹⁰, Catherine Faron-Zucker⁹, Jean-Paul Stromboni⁹ et Alain Giboin¹⁰ est un outil qui sert aux étudiants lors des travaux dirigés. Pour le professeur, cet outil permet de mettre son cours sous forme électronique et d'en faciliter la navigation. QBLS permet d'accéder au cours et de tester les connaissances des élèves. En effet cette application propose des questions sur le cours et propose des liens vers des connaissances utiles pour répondre à la question. L'élève, qui utilise l'outil, doit aller trouver la réponse par lui-même.

Avec les technologies sémantiques, les documents sont mis sous forme électronique en les structurant en fiche d'informations d'où ressortent des mots et connaissances clés.

Mathis

Le projet Mathis (MATHématique Information web Service) est un projet initialisé par l'équipe Mainline du laboratoire I3S. Le but de ce projet est de développer des outils pour annoter et

¹⁰Équipe Mainline du laboratoire I3S

¹¹Acacia, inria

rechercher des documents scientifiques (plus particulièrement mathématiques) sur le web. Ce projet se découpe en trois applications :

- Mathis formula publisher : éditeur de formules mathématiques annotées grâce aux ontologies pour décrire sémantiquement les domaines concernés
- Mathis annoter : éditeur d'annotations
- Mathis search engine : moteur de recherche qui exploite les annotations créées avec les outils précédents

Les wiki sémantiques

Un wiki est un logiciel qui permet de gérer le contenu de sites web. Les pages web du site sont alors modifiables par des utilisateurs spécifiques. C'est un programme qui facilite le travail collectif. Ces espaces contiennent essentiellement des textes et des hyperliens.

Les wiki sémantiques sont donc des wiki que l'on peut modifier en utilisant les outils du web sémantique. Sur le même principe que le wiki « classique », le wiki sémantique permet de créer une base de données dont tous les objets sont liés entre eux par différentes relations explicitées. Les pages d'un wiki sémantique sont donc organisées de façon complexe et sont reliées les unes avec les autres. L'utilisation des techniques sémantiques permet donc de définir et d'identifier chaque mot et surtout d'être interprété par des machines.

Nous allons vous présenter quelques exemples de wiki sémantiques.

Freebase

Première réalisation de l'entreprise Metaweb créée par Dannys Hillis et Robert Cook qui a pour but, selon eux, de créer « *une base de connaissances communes, une base de données structurée, interrogeable, constituée et modifiée par une communauté de contributeurs* », ou encore « *un espace public des données* ».

Ils ont voulu créer une base de données structurées, intelligibles et exploitable autant par des humains que des machines.

Les données y sont décrites sous la forme de champs. Chaque information a un rôle particulier selon l'espace qu'elle occupe. Freebase peut aussi être complété par les utilisateurs. Ces derniers sont aidés par des outils simples et des listes pré-établies pour remplir les champs des différentes fiches et organiser les documents. Les utilisateurs sont donc invités à compléter les informations et à aider à l'indexation. Ils ont la possibilité de tout modifier (sous contrôle) : les ontologies, les champs de données, etc. Les données sont aussi structurées en taxonomie thématiques.

Cependant Freebase n'utilise pas la technologie du web sémantique. Il se sert du langage MQL (Metaweb query langage) qui est un langage créé par Metaweb. Or le principe d'un wiki sémantique est l'utilisation de standard pour décrire les données (RDF) et d'un même langage de requête (SPARQL).

Semantic MediaWiki

Semantic MediaWiki est une extension de MediaWiki. MediaWiki est un logiciel libre, développé initialement pour wikipédia. Il est utilisé actuellement par de nombreux projets et sites utilisant les propriétés du wiki. Cette extension permet de la recherche, l'organisation, le balisage, la navigation et le partage du contenu de différents wiki. C'est donc l'apport des technologies du web sémantique au wiki. Elle introduit les annotations des articles des wikis en ajoutant deux propriétés : les relations qui permettent de décrire les liens entre deux articles et les attributs qui attribuent à un article d'autres types de données : chaînes de caractères, nombres, etc.

Semantic MediaWiki permet, grâce aux balisages, de rendre les contenus des wikis utilisables par des ordinateurs.

DBpedia

DBpedia est un projet mené par deux chercheurs allemands. Le but de ce projet est d'extraire des informations de Wikipédia afin de les rendre disponible dans un format utilisable dans des entrepôts de données avec des requêtes compliquées.

DBpedia va chercher dans les bases de données relationnelles de Wikipédia, les données contenues dans les « infobox » (box présentes à droite d'un article Wikipédia et contenant des informations normalisées). Ces données sont ensuite décrites en RDF/XML suivant une ontologie définie pour chaque type d'objet. Elles sont ensuite conservées dans un entrepôt spécialisé et peuvent être interrogées en SPARQL.

Les sciences du vivant

Les nouvelles techniques, utilisées dans le domaine de la biologie moléculaire comme le séquençage des génomes ou les puces à ADN, produisent une grande masse de données rendent le travail des biologistes plus compliqué. Ce domaine est en constante progression et chaque nouvelle découverte implique de plus en plus de données. L'accès à ces données tout comme leurs interprétations et annotations sont des tâches délicates à réaliser avec les outils classiques du biologiste.

En plus de l'énorme volumétrie des données, il existe de nombreuses terminologies employées par les biologistes : chacune étant établie pour un sous-domaine précis de la biologie. Deux sous-domaines différents peuvent avoir un même terme commun dont la signification sera différente.

« Par ailleurs, un grand nombre de bases de données spécialisées voient le jour et sont accessibles à tous sur la toile. Cependant elles utilisent généralement des entrées différentes, mais aussi des terminologies propres à chacune. Ainsi, les biologistes ne peuvent avoir une vision globale d'une connaissance disponible à partir d'une seule et même source car il n'existe pas de base de données qui mutualise toutes celles disponibles en biologie. »¹²

¹² « Enrichissement automatique d'une base de connaissances biologiques à l'aide des outils du web sémantique » Ines Jilani, Florence Amardeilh

Le langage naturel que l'on trouve sur le web est trop ambigu. Les technologies du web sémantique sont donc ici précieuses pour expliciter les contenus du web. La biologie utilise les ontologies afin de proposer un ensemble structuré, une description de la structure des informations disponibles sur le sujet. C'est un modèle représentatif de l'ensemble des concepts dans le sous domaine de la biologie concerné et de leur relations. L'ontologie permet de définir les concepts d'un domaine en les organisant hiérarchiquement et en définissant leurs propriétés.

Les outils du web sémantiques peuvent permettre d'enrichir automatiquement des bases de données.

« La mise en oeuvre du peuplement de telles ontologies grâce aux solutions proposées par le Web sémantique passe par le traitement du langage naturel, complété par une ontologie de référence. Dans ce cas, la tâche consiste à repérer dans le texte les instances, existantes ou nouvelles, de cette ontologie. Lorsque de nouvelles instances sont identifiées dans le texte, extraites puis reliées à l'ontologie, il s'agit alors de peupler cette ontologie, c.-à-d. d'enrichir la base de connaissances y étant associée avec ces nouvelles instances. »¹³

Intelligence Artificielle

Le but du web sémantique et les technologies utilisées nous mène à introduire le concept d'Intelligence Artificielle (IA). L'IA est une « *Discipline scientifique relative au traitement des connaissances et au raisonnement, dans le but de permettre à une machine d'exécuter des fonctions normalement associées à l'intelligence humaine : compréhension, raisonnement, dialogue, adaptation, apprentissage, etc.* »¹⁴. Ou bien encore la « *la science dont le but est de faire faire par une machine des tâches que l'homme accomplit en utilisant son intelligence.* »¹⁵.

Histoire de l'IA

Depuis l'Antiquité, les hommes rêvent d'une machine intelligente. Ce ne fut qu'en 1950 que l'idée de l'intelligence artificielle apparaît dans l'article « Computing Machinery and Intelligence » d'Alain Turing. Il y présente une expérience, le test de Turing, qui tente la définition d'un standard permettant de qualifier une machine de « consciente ».

L'Intelligence Artificielle apparaît en 1956 lors d'une rencontre à Dartmouth College entre John McCarthy et Marvin Minsky. Cette rencontre est le résultat d'une effervescence autour des questions liées d'une part à la possibilité de construire des "machines à penser", voire des "machines pensantes" et d'autre part à la comparaison du fonctionnement du cerveau humain avec les premiers ordinateurs qui venaient d'apparaître.

Ces deux chercheurs n'ont pourtant pas la même appréhension de cette discipline : John McCarthy prône une vision purement logique de la représentation des connaissances alors que Marvin Minsky privilégie l'usage de la représentation structurées, de schémas. L'expression Intelligence Artificielle fut introduite par McCarthy.

¹³ « Enrichissement automatique d'une base de connaissances biologiques à l'aide des outils du web sémantique » Ines Jilani, Florence Amardeilh

¹⁴ http://www.futura-sciences.com/fr/definition/t/informatique-3/d/intelligence-artificielle_555/

¹⁵ <http://www.math-info.univ-paris5.fr/~pastre/IA>

La naissance de l'IA a été largement influencée par leurs contemporains : Warren McCulloch et Walter Pitts en neurophysiologie avec les premiers modèles de réseaux de neurones artificiels, Norbert Wiener en cybernétique (sciences des mécanismes de communication par l'étude de l'information et des principes d'interaction ou « encore la science des analogies maîtrisées entre organismes et machines »), Claude Shannon en théorie de l'information, John von Neumann avec l'architecture des calculateurs, Alan Turing avec les fonctions calculables par machine.

Puis viennent les premiers programmes d'IA. Newell, Simon et Shaw proposent en 1956 un premier programme de démonstration de théorème en logique, puis un « résolveur de problème général » basé sur l'évaluation de la différence entre la situation à laquelle le résolveur est arrivé et le but qu'il a à atteindre.

Très rapidement l'IA s'intéresse au développement de programmes capables de jouer aux échecs (l'idée avait été présentée en 1950 par Shannon). Les premiers programmes apparaissent dans les années 60 et au fur et à mesure qu'ils sont perfectionnés, ils battent des joueurs de niveaux de plus en plus élevés. En 1997 eu lieu la première victoire d'un ordinateur sur un maître des échecs : l'ordinateur « Deep Blue » bat Gary Kasparov.

La programmation d'intelligence artificielle se poursuit en 1963 avec le développement d'un test « d'intelligence » par Evans. Ce test permet de trouver par analogie la quatrième figure géométrique sur une série de trois figures. Par la suite David Waltz en 1975 permet la résolution de problèmes par propagation de contraintes.

Le traitement de textes ou de dialogue en langue naturelle (compréhension ou production automatique) est un sujet qui a très rapidement préoccupé les chercheurs. Un système de dialogue en langue naturelle fut mis au point en 1965 sous le nom de Eliza par Weisenbaum. Eliza était capable de repérer des expressions clés dans une phrase et d'en reconstruire à partir des expressions sélectionnées. Les utilisateurs étaient piégés un moment, croyant parler à un interlocuteur humain, avant de se rendre compte qu'ils dialoguaient avec un ordinateur. Il ne construisait aucune représentation des phrases du dialogue et n'avait aucune compréhension. Le premier à reproduire des représentations et à les exploiter dans des dialogues fut Shrdlu mis au point par Terry Winograd en 1971.

Les années 70 et 80 connaissent un engouement pour le développement de « systèmes experts ». Nous citerons Dendral en chimie, Mycin en médecine, Hearsay II en compréhension de la parole, ou encore Prospector en géologie. Ces systèmes ont une connaissance experte dans un domaine spécialisé qui est exprimée avec des règles et contraintes de la forme "si...alors...". Nous en parlerons plus en détails ci-dessous.

Les années 70 sont aussi des années consacrées aux expérimentations des premiers robots mobiles.

L'intelligence artificielle a connu un premier développement rapide dès l'apparition des premiers ordinateurs scientifiques. Cependant elle a été ralenti par le manque de capacité des composants électroniques de cette époque. Ce n'est qu'au début des années 1990 que les progrès ont repris avec la généralisation des ordinateurs de bureau, des réseaux de télécommunications et des matériels incorporant de plus en plus de solutions IA liées à des besoins spécifiques.

On distingue deux types d'intelligences artificielles : l'IA faible et l'IA forte. Les deux se distinguent par la nature de l'intelligence mis en jeu.

L'IA faible

L'IA faible est une approche pragmatique de l'intelligence artificielle. Elle cherche à construire des systèmes autonomes et des algorithmes capables de trouver des solutions similaires à celles des humains pour un certain type de problème. C'est donc la création de machines qui simulent l'intelligence humaine. Les programmes/machines ainsi obtenus ne sont pas dotés d'intelligence, ni de capacités de raisonnement ou d'évolution qui leur seraient propres. Il s'agit d'une reconstitution de l'intelligence humaine par la programmation d'apprentissages. On peut parler d'une continuité de la recherche opérationnelle, de l'aide à la décision et du data mining.

L'IA faible s'appuie sur les réseaux neuromimétiques, systèmes qui imitent les neurones vivants. Comme les neurones, l'IA faible s'appuie sur la reconnaissance de situations ou la similarité d'une situation avec une autre déjà rencontrée pour résoudre le problème posé. C'est une forme d'intelligence artificielle qui est rapide et robuste. Les applications de l'IA faible sont nombreuses dans la domotique, comme dans le traitement du langage (reconnaissance vocale, écrite...), etc.

L'IA forte

L'intelligence artificielle forte fait référence aux machines capables de produire un comportement intelligent, d'éprouver une impression de réelle conscience de soi, de « vrais *sentiments* » et « *une compréhension de ses propres raisonnements* »¹⁶. On attribue à cette intelligence en plus de la reproduction de l'intelligence humaine, tous les concepts abstraits qui en découlent tels que la conscience des actes ou la compréhension de ces derniers.

L'IA forte a pour objectif de reproduire le cerveau humain. La conscience humaine et les raisonnements humains sont créés par un support biologique (le cerveau) qui peut être considéré comme un support matériel. Les scientifiques, dans le domaine de l'IA forte, aspirent à créer un jour une intelligence consciente sur un support matériel non biologique. Selon eux, si nous n'avons toujours pas réussi à créer des ordinateurs ou robot aussi intelligent que les humains, c'est à cause d'un problème de conception et non pas d'un problème matériel.

¹⁶ Dictionnaire de l'informatique (1975), André Le Garff



L'IA et la représentation de la conscience reposent sur deux approches : le cognitivisme et le connexionnisme.

Le cognitivisme est basé sur l'hypothèse que la pensée humaine est un processus de traitement de l'information. Il considère que le vivant fonctionne comme un ordinateur, qu'il manipule des symboles élémentaires. Le connexionnisme décrit quand à lui des « *phénomènes mentaux comme des processus émergents de réseaux d'unités simples interconnectées* »¹⁷. L'intelligence serait donc comme un système en réseau de multiples unités simples.

L'intelligence artificielle est présente dans nos vies par les jeux vidéos, les aides à l'analyse décisionnelle, etc. Mais elle soulève aussi de nombreuses questions. Les questions les plus courantes sont les suivantes : l'Homme parviendra-t-il à créer une machine (un robot) aussi intelligente que lui ? Si oui, que se passerait-il si les machines devenaient plus intelligentes que lui ? De là viennent toutes les hypothèses et les scénarios qui décrivent les machines parvenant à nous dominer et à nous exterminer.

Certains grands projets de l'IA utilisent les technologies du web sémantique. C'est le cas du projet Cyc. Lancé par Douglas Lenat, de la société Microelectronics and Computer Technology Corporation, le projet Cyc a pour but de créer un grand puit de connaissances qui pourrait être utilisé par des applications d'IA pour raisonner comme un être humain. Pour ça il tente de développer une ontologie et une base de données de la connaissance générale.

L'IA est donc un domaine d'application du web sémantique. Cependant, le web sémantique n'est qu'à ces débuts et l'IA n'évolue pas de manière significative. Les projets alliant les deux ne seront pas les grands projets de demain.

¹⁷ <http://www.maxime-viry.com/blog/intelligence-artificielle-evolution-revolution>

Systemes experts et moteurs d'inférence

Un système expert est une application qui peut effectuer les raisonnements logiques ou reproduire les mécanismes cognitifs d'un expert d'un domaine précis. C'est un outil d'aide à la décision capable de répondre à des questions par des raisonnements basés sur des faits et règles connus. Il est composé d'une base de connaissance : (bases de faits, bases de règles) et d'un moteur d'inférence. Le système expert va tenter de représenter l'expérience de l'expert c'est-à-dire ses connaissances pratiques qui ne peuvent pas se trouver dans des documents théoriques.

Le moteur d'inférence est un logiciel qui réalise les déductions logiques du système expert en utilisant les deux bases de celui-ci. Il déduit des faits à partir d'autres faits. Il existe trois types de moteurs d'inférence : les moteurs à chaînage avant, les moteurs à chaînage arrière et les moteurs mixtes. Les moteurs à chaînage avant déroulent leur logique en partant de la base de connaissance pour approcher les faits recherchés par le problème. Les moteurs à chaînage arrière partent des faits recherchés par le problème pour remonter jusqu'aux faits connus via les règles existantes. Les moteurs mixtes utilisent les deux approches précédentes.

Les moteurs d'inférence s'appuient la logique formelle (issue des logiques mathématiques) pour résoudre le problème donné. La logique formelle se base sur des connecteurs tels que la disjonction (« ou » logique), la négation, l'équivalence, etc et des prédicats. Par exemple, voici un type de règle fréquemment utilisée : Si A implique B et que A est vraie, alors B est vraie.

Il existe de nombreux moteurs d'inférence. Un des plus connu est le moteur Mycin, développé à partir de 1972 pour le domaine médicale. Il permet une aide au diagnostic et au traitement d'infections bactériennes. Ce programme a été testé en 1979 face à 8 médecins sur 10 cas réels et il a gagné. Ce logiciel n'a pas été très utilisé du fait de sa complexité, de sa mauvaise interface et du fait qu'il était difficile d'évaluer ses résultats. Cependant, il a montré aux hommes que l'on peut approcher un domaine d'expertise avec un programme.

Pour mettre en place ce type de système d'information, il faut déterminer à l'avance les règles de prise de décision. Il faut donc disposer d'une terminologie complète du domaine étudié et de règles d'aide à la décision, confrontées régulièrement aux réalités. Les ontologies sont donc des outils appropriés à la création de système expert.

4. Perspectives critiques

Le débat sur l'ontologie existe depuis longtemps, et continue en philosophie. N'est-il pas inévitable qu'il existe et persiste sur le web sémantique ?

Avec ses militants inconditionnels et ses pourfendeurs ou détracteurs, le Web Sémantique ouvre un débat qui porte sur le caractère utopique ou non de la démarche. Certains pensent que c'est une idée géniale, qu'il faut enrichir, faire avancer et qui, de toute façon, finira par s'imposer. D'autres pensent que pour des raisons psycho-sociologiques, sociétales,

politiques, commerciales, le web sémantique ne peut pas se généraliser car il demande une trop grande discipline.

Les discours utopiques

Internet regorge de discours utopiques sur le web sémantique. On ne peut pas faire de recherche sur ce sujet sans tomber sur de nombreux blogs faisant l'apologie du web qu'ils appellent le web 3.0.

Ces discours sont-ils réalistes ou utopiques ? C'est la question que l'on peut se poser en écoutant les discours de Tim Berners Lee sur le sujet. Nous vous proposons un extrait d'une interview de Tim Berners Lee dans le *Courier Unesco*¹⁸. A la question « *Pouvez-vous nous parler du «Web sémantique» auquel vous travaillez actuellement?* » il répond :

« J'ai un double rêve pour le Web. D'une part, je le vois devenir un moyen très puissant de coopération entre les êtres humains. Et dans un second temps, j'aimerais que ce soit les ordinateurs qui coopèrent. Les machines deviennent capables d'analyser toutes les données circulant sur le réseau: contenus, liens et transactions entre les personnes et les ordinateurs. Le Web sémantique ira chercher l'information dans diverses bases de données, aussi bien dans des catalogues en ligne que sur des sites météo ou boursiers, et il permettra que toute cette information soit traitée par les ordinateurs. Ce n'est pas possible dans l'état actuel des choses parce que les données en ligne ne sont pas compatibles ni formatées de manière à être directement analysées par des machines. Les pages Web sont uniquement destinées à la lecture humaine.

Le Web sémantique répondra aussi aux vœux de tous ceux qui voudraient disposer d'un moteur de recherche solide. Pour l'instant, les moteurs de recherche vous livrent des milliers de pages en réponse à une seule demande. Or, il est impossible d'étudier le contenu de toutes ces pages. Avec le Web sémantique, le moteur de recherche répondra: «Voilà un objet dont je peux mathématiquement garantir qu'il répond au critère souhaité». En bref, les moteurs de recherche deviendront plus fiables et plus efficaces. Quand mon rêve sera réalisé, le Web sera un univers où la fantaisie de l'être humain et la logique de la machine pourront coexister pour former un mélange idéal et puissant. »

Tout d'abord l'utilisation du terme « rêve » pour parler du web sémantique nous amène à penser que l'image qu'il nous présente est utopique. Si on regarde aussi l'article qu'il a co-écrit en 2001, « the semantic web », on s'aperçoit qu'il conçoit l'utilisation du futur web comme étendue dans les pratiques quotidiennes, avec une certaine banalisation de la toile.

Outre le fait de que le travail paraît immense pour passer du web actuel au web sémantique, il faut aussi voir que ce web se base sur le sens et que dans le monde entier, le sens diffère et dépend des cultures. En effet, les facteurs culturels définissent notre vision du monde et notre appréhension aux choses, s'ils devaient être pris en compte, un mot ne pourrait alors pas se résumer à une même définition.

¹⁸ http://www.unesco.org/courier/2000_09/fr/dires.htm

Les adeptes du web sémantique soulignent le fait que son apport serait immense. Il offrirait de nombreuses possibilités pour les outils de recherche qui, à l'heure actuelle, ne sont pas des plus pratiques. En effet, le web est une bibliothèque gigantesque regorgeant d'informations qui peuvent être utiles, mais qui sont le plus souvent inutiles. Lors d'une recherche sur internet, on passe plus de temps à lire des articles qui ne vont pas nous intéresser qu'à comprendre et analyser des résultats pertinents. Le web sémantique permettrait donc des progrès dans la recherche scientifique et faciliterait les échanges de connaissance. Tout comme il faciliterait le traitement automatique des données disponibles sur internet.

De plus, ils soulignent le fait que cette avancée s'appuie sur l'infrastructure déjà existante du web et qu'il ne s'agirait que d'ajouter des protocoles. Pour les utilisateurs aucune différence ne sera visible, il ne sera pas possible de différencier une page dont le contenu a la structure du web sémantique d'un site dont le contenu est classique.

Il est clair que les apports du web sémantique ne s'arrêtent pas au web. Il permettrait de faire communiquer les machines entre elles. Il fera donc partie intégrante de notre vie.

Les discours critiques

La plupart des critiques du web sémantique reprochent à ses défenseurs d'en faire un portrait trop utopique.

Pour Luciano Floridi le web sémantique est trop ambitieux. D'après lui il «*repose sur une version forte du projet de l'Intelligence Artificielle [IA], et, par conséquent, il est techniquement irréalisable.*»

Parmi les arguments des opposants vient celui-ci : le coût du web sémantique apparaît à certain comme un obstacle. Le travail nécessaire pour enrichir un contenu avec les outils du web sémantique n'est pas négligeable.

Mais les arguments les plus cités sont que les producteurs de contenu ne sont pas toujours altruistes et que le concept du web sémantique n'est pas accessible à tout le monde. En effet, les motivations de la mise en ligne de différents contenus sont la plupart du temps économiques voire politique. Si les raisons sont économiques, le méta-données peuvent être détournées très facilement afin de tricher sur le contenu d'un document et peuvent servir au spam. La même technique peut être utilisée afin de diffuser des discours ou théories politiques.

L'article virulent de Cory Doctorow «*Putting the torch to seven straw-men of the meta-utopia*», traduit de l'anglais par Clément Pillias sous le titre «*les sept illusions de la méta-utopie*» énonce de façon très pessimiste sept traits humains qui, selon lui, empêcheront l'essor du web sémantique.

Ces illusions sont les suivantes :

- Les gens mentent : les méta-données seront truquées à des fins commerciales, politiques...
- Les gens sont paresseux : donc les gens ne rempliront pas, ou pas totalement les méta-données
- Les gens sont stupides : les gens ne sont pas assez attentifs pour ne pas faire d'erreur et ils auront du mal à s'approprier ce concept
- Mission : Impossible — connais-toi toi-même : difficulté d'observation et de description
- Les schémas ne sont pas neutres
- Les métriques influencent les résultats : difficulté de s'entendre sur une métrique à cause des différents intérêts liés
- Il y a plus d'une façon de décrire quelque-chose

Cet article est assez fort et dénonce un caractère humain qui selon l'auteur, serait bloquant pour une bonne utilisation du web sémantique.

Nous citerons un dernier exemple de critique qui s'attaque plus particulièrement aux ontologies. Clay Shirky dans son article « L'Ontologie Est Surfaite : Catégories, Tags et Liens » critique les différentes manières d'organiser des données : *« Dans les systèmes de classification de travail, le succès n'est pas "Avons-nous obtenu le rangement idéal ?" mais plutôt "A quel niveau de proximité sommes-nous, et sur quelles mesures ?" L'idée d'un schéma parfait est simplement un idéal Platonique. Néanmoins, je veux avancer que même l'idéal ontologique est une erreur. Utiliser même la perfection théorique comme une mesure pratique du succès mène à une mauvaise application des ressources. »*

Il commence par nous citer un exemple qui est la classification périodique des éléments pour nous montrer que même ce classement, qui est une référence scientifique, est faussé. En effet, l'hélium y est classé gaz noble à l'époque de la création de ce tableau alors que depuis des scientifiques ont montré que, pour une certaine température, l'hélium n'est plus un gaz.

La classification ne peut pas être parfaite : elle dépend de la conception du monde de chacun, et chaque chose ne peut pas forcément être classé.

Pour lui, l'ontologie est une des meilleures solutions mais elle ne sera pas une représentation parfaite de la réalité. Déjà, on ne peut pas représenter le « Tout » avec une ontologie mais des parties du « Tout ». De plus, *« si vous avez des utilisateurs naïfs, si vos catalogueurs ne sont pas des experts, s'il n'y a personne pour dire officiellement ce qui se passe, l'ontologie va être une mauvaise stratégie. »*

Tous s'accordent à dire que l'idée est bonne, que les apports seraient intéressants. Mais pour les pourfendeurs c'est un projet utopique à plusieurs niveaux : à un niveau technique comme à un niveau social, politique et commercial. Tout d'abord pour eux les difficultés techniques sont trop grandes. Ensuite, l'humanité est telle que cet outil serait détourné de son but premier et exploité à mauvais escient. Beaucoup en concluent que c'est un projet long terme utilisable pour la communication entre les machines.

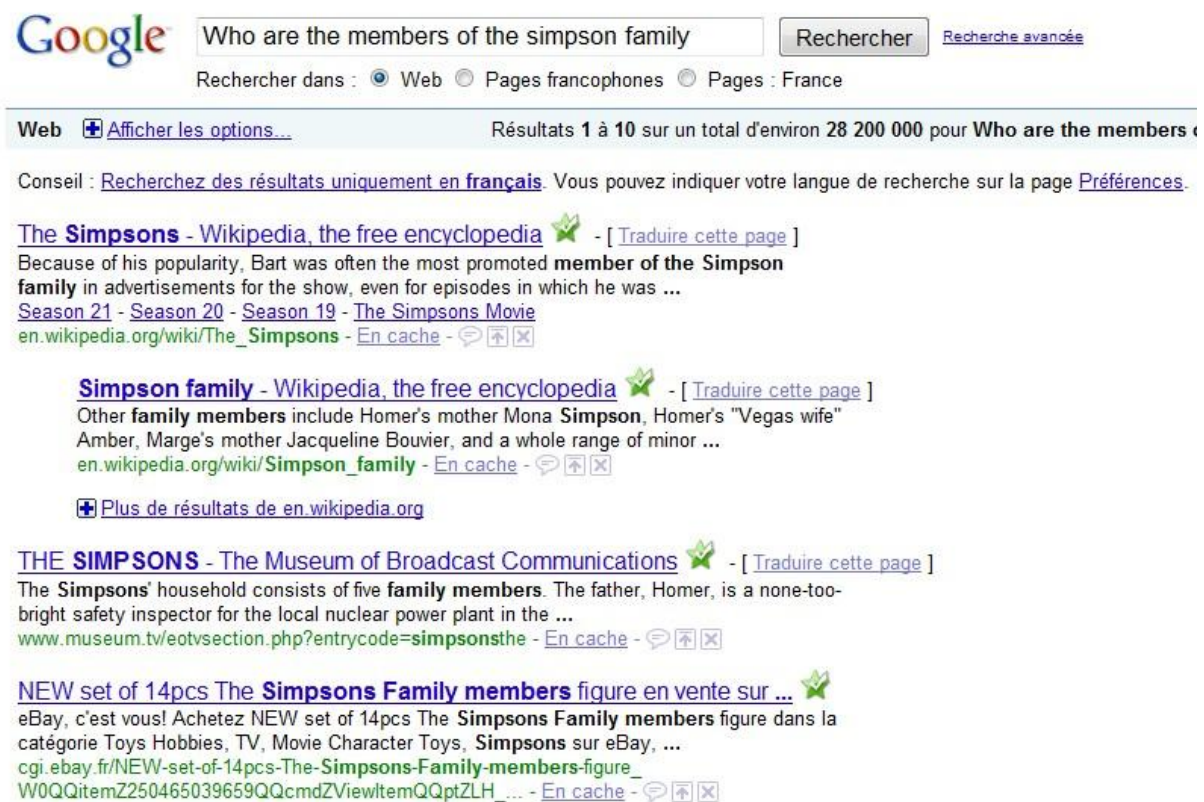
Partie plus pragmatique

1. Moteur de recherche sémantique Powerset

Nous allons vous présenter Powerset qui est un moteur de recherche s'approchant d'un moteur de recherche sémantique. Lorsqu'un moteur de recherche traditionnel recherche une concordance entre la requête de l'internaute et les données de son index, le moteur sémantique tente de répondre à la question qu'on lui pose. Les moteurs de recherche actuels tels que Google fonctionnent par mots clés. Si on tape dans Google une question, il va chercher, dans son index, des pages contenant la question (mot pour mot), et nous donnera les résultats contenant les mots de la question (mis à part si quelqu'un a écrit sur une page la question que l'on se pose). Le moteur sémantique tente de formuler une réponse en fonction de la question posée.

Actuellement, Powerset ne fonctionne qu'avec Wikipédia.

Nous allons vous faire une comparaison entre une recherche sur Google et une recherche sur Powerset. Pour cela, nous avons choisi une question que nous allons soumettre aux deux moteurs de recherche : Who are the members of the simpson family ? Voici le résultat obtenu avec Google :



The screenshot shows a Google search interface. The search bar contains the text "Who are the members of the simpson family". To the right of the search bar is a "Rechercher" button and a link to "Recherche avancée". Below the search bar, there are radio buttons for "Web" (selected), "Pages francophones", and "Pages : France". Below this, there is a section for "Web" with a link to "Afficher les options...". The search results are displayed below, showing "Résultats 1 à 10 sur un total d'environ 28 200 000 pour Who are the members of the simpson family". A "Conseil" message suggests searching in French. The first result is "The Simpsons - Wikipedia, the free encyclopedia" with a green star icon and a link to "Traduire cette page". The snippet for this result reads: "Because of his popularity, Bart was often the most promoted member of the Simpson family in advertisements for the show, even for episodes in which he was ...". Below this are links for "Season 21", "Season 20", "Season 19", and "The Simpsons Movie". The second result is "Simpson family - Wikipedia, the free encyclopedia" with a green star icon and a link to "Traduire cette page". The snippet reads: "Other family members include Homer's mother Mona Simpson, Homer's 'Vegas wife' Amber, Marge's mother Jacqueline Bouvier, and a whole range of minor ...". Below this is a link to "Plus de résultats de en.wikipedia.org". The third result is "THE SIMPSONS - The Museum of Broadcast Communications" with a green star icon and a link to "Traduire cette page". The snippet reads: "The Simpsons' household consists of five family members. The father, Homer, is a none-too-bright safety inspector for the local nuclear power plant in the ...". The fourth result is "NEW set of 14pcs The Simpsons Family members figure en vente sur ..." with a green star icon. The snippet reads: "eBay, c'est vous! Achetez NEW set of 14pcs The Simpsons Family members figure dans la catégorie Toys Hobbies, TV, Movie Character Toys, Simpsons sur eBay, ...".

Voici les résultats obtenus avec Powerset :

Powerset

Wikipedia Articles

Who are the members of the simpson family

Wikipedia Articles

- Simpson family** Homer Jay **Simpson**, voiced by Dan Castellaneta, is the father of the **Simpson family**. ... In gratitude, he bought gifts for each **member of the family** and paid Homer back with a vibrating chair, along with his forgiveness.
- Bart Simpson** Bartholomew JoJo "**Bart**" **Simpson** is a fictional main character in the animated television series The Simpsons and part of the eponymous **family**. ... His lack of smarts can also be attributed to the hereditary "**Simpson Gene**", which affects the intelligence of most male **members of the Simpson family**.
- Lisa Simpson** **Lisa Marie Simpson** is a fictional main character in the American animated television series The Simpsons. ... On April 9, 2009, the United States Postal Service unveiled a series of five 44 cent stamps featuring Lisa and the four other **members of the Simpson family**.
- The Simpsons** | Main cast **members** | ... Because of his popularity, **Bart** was often the most promoted **member of the Simpson family** in advertisements for the show, even for episodes in which he was not involved in the main plot.
- Maggie Simpson** **Margaret "Maggie" Simpson** is a fictional character in the animated television series The Simpsons. ... On April 9, 2009, the United States Postal Service unveiled a series of five 44 cent stamps featuring Maggie and the four other **members of the Simpson family**.
- Homer Simpson** **Homer Jay Simpson** is a fictional main character in the animated television series The Simpsons and father of the eponymous **family**. ... On April 9, 2009, the United States

Comme on peut voir ci-dessus, Powerset surligne en orange clair la réponse à la question posée. Il surligne aussi les mots de la question en jaune. L'avantage de Powerset sur Google (dans cet exemple) est le suivant : en plus de trouver une page traitant de la thématique sur laquelle on s'interroge, ce moteur sélectionne le paragraphe répondant à la question que l'on se pose.

2. True Knowledge

True knowledge n'est pas un moteur de recherche mais un « moteur de questions ». Il utilise Wikipédia et sa propre base de données sémantique afin de répondre aux questions. Les requêtes envoyées par l'utilisateur doivent être sous forme de questions et comporter un « ? ».

C'est un plugin que l'on peut ajouter à Firefox.

Nous vous proposons deux questions que l'on a posé dans Google, avec le plugin True knowledge.

Voici la réponse à la question : What is the difference between twelve and nine?



What is the difference between twelve and nine?

Rechercher

[Recherche avancée](#)

Rechercher dans : Web Pages francophones Pages : France

Web [Afficher les options...](#)

Résultats 1 à 10 sur un total d'environ 152 000 000 pour What is the difference between

3

Good answer? [Yes](#) [No](#) source: True Knowledge

Conseil : [Recherchez des résultats uniquement en français](#). Vous pouvez indiquer votre langue de recherche sur la page [Préférences](#).

[What is the difference between twelve and nine? - True Knowledge](#) - [[Traduire cette page](#)]

what is the name of the ... **what is the difference between twelve and nine**. how ... What is the name of the pass between Pakistan and Afghanistan near ...

www.trueknowledge.com/.../what_is_the_difference_between_twelve_and_nine -

En cache -

[12-hour clock - Wikipedia, the free encyclopedia](#) - [[Traduire cette page](#)]

... confusion about the **difference between** midnight at the start and end of a as oh one to oh **nine** (nought or zero can also be used instead of oh). ...

en.wikipedia.org/wiki/12-hour_clock - En cache - [Pages similaires](#) -

[Moser Rearend And The Difference Between 12 Bolts And 9 Inch Photo](#) - [[Traduire cette page](#)]

View this Moser Rearend And The **Difference Between** 12 Bolts And 9 Inch photos. Get the latest information on driveline innovations from Rob Moser of Moser ...

www.chevyhiperformance.com/.../photo_09.html - En cache -

[Difference between Jaguar XJR-9, 10, 11, 12 - Car Forums and ...](#) - [[Traduire cette page](#)]

4 messages - 3 auteurs - Dernier message : 3 oct.

Does anybody how to spot the **difference between** these cars? ... The difference in engine capacities between the IMSA and Group C cars had to ...

www.automotiveforums.com/t629339-difference_between_jaguar_xjr_9_10_11_12.html -

En cache -

Terminé

La réponse de True knowledge (TK) apparaît en haut, au dessus des résultats de Google. On voit bien que TK comprend la question et y répond en termes mathématiques, alors que Google ne donnera que des résultats contenant les mots de la question.

Nous avons ensuite posée la question suivante : what does API stand for?



what does API stand for?

Rechercher

[Recherche avancée](#)

Rechercher dans : Web Pages francophones Pages : France

Web [Afficher les options...](#)

Résultats 1 à 10 sur un total d'environ 10 600 000 pour what

["Application Programming Interface"](#)

Good answer? [Yes](#) [No](#)

source: [TrueKnowledge](#)

[Définitions du Web pour API](#)



La palatalisation est une modification phonétique dans laquelle un son est produit par une partie plus à l'avant du palais dur que celle utilisée pour le son d'origine.

fr.wikipedia.org/wiki/API - [Définition dans son contexte](#)

[API - What does API stand for? Acronyms and abbreviations by the ...](#) - [[Traduire cette page](#)]

Acronym, Definition. API, Application Programming Interface. API, Academic Performance Index. API, Application Program Interface ...

acronyms.thefreedictionary.com/API - [En cache](#) - [Pages similaires](#) -

[What does API stand for?](#) - [[Traduire cette page](#)]

What does API stand for? We've got 47 definitions for API in the Abbreviations.com acronyms and abbreviations directory.

www.abbreviations.com > ... > Software - [En cache](#) - [Pages similaires](#) -

[API - Definition by AcronymFinder](#) - [[Traduire cette page](#)]

Acronym Finder: Definition of API. What does API stand for? ... One-Stop Shop for API Standards Purchase and Download API Documents. global.ihs.com ...

www.acronymfinder.com/API.html - [Pages similaires](#) -

[What does "api" stand for? - True Knowledge](#) - [[Traduire cette page](#)]

A Partner in Technology (API) is a maine consulting firm providing technology solutions. ...

What does API stand for? ... this is what API Technology wants

On peut voir ici que cette question est régulièrement posée sur internet. C'est pourquoi Google nous présente des sites contenant la question entière. Cependant, si on veut avoir la réponse à notre question, il faut que nous ouvrons chaque lien jusqu'à ce que l'on trouve la définition de cet acronyme. TK lui nous donne directement la réponse à la question posée.

Cependant TK ne répond pas à toutes les questions qu'on lui pose. Si nous posons la même question que pour Powerset : Who are the members of the simpson family ?, nous n'obtenons aucun résultat.

3. Cycle de vie du web sémantique

Pour tenter de définir où en est le web sémantique, nous nous sommes tournés vers une entreprise spécialisée dans l'étude des technologies : Gartner Group.

Hype Cycles

Un hype cycle est un graphique représentant la maturité, l'adoption et l'économie des nouvelles technologies.

Voici l'explication d'un Hype cycle traduit de l'anglais par Ludovic Bour, professionnel de la fonction marketing dans le domaine des services aux entreprises :

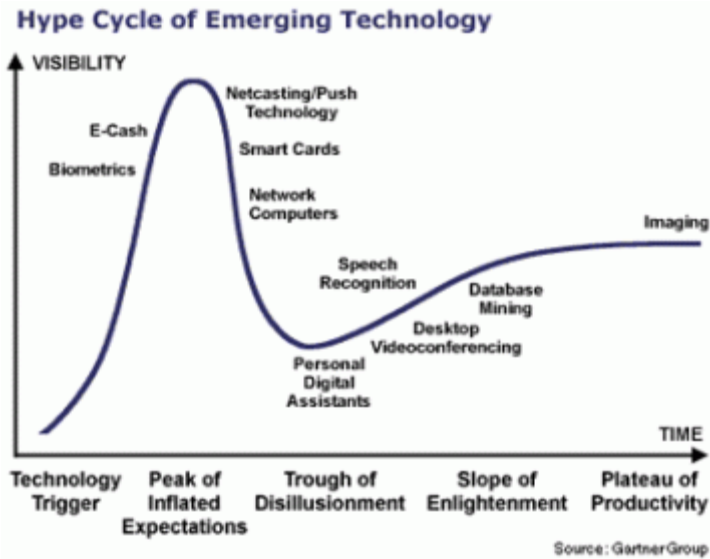
« Ce schéma pour la compréhension des cycles de vie des produits technologiques (IT) permet de suivre les principales phases de l'adoption d'une technologie par le marché, de la création jusqu'à la maturité.

L'intérêt de ce schéma est qu'il permet de faire la distinction entre les phases d'"euphories" et les phases d'adoption réelle de la technologie. Au final ce schéma permet aux décideurs d'entreprise de déterminer si le moment est venu d'adopter une technologie ou non.

5 phases sont identifiées dans l'émergence d'une technologie :

- 1. le démarrage Technologique (présentation publique, lancement produit, évènements générant un intérêt important dans la presse ou sur le marché)*
- 2. le sommet des espérances gonflées (durant cette phase d'enthousiasme extrême et de projection irréaliste, une rafale d'activités publicitaires menées par des leaders technologiques générant quelques succès mais beaucoup d'échecs car la technologie est poussée jusqu'à ses limites. Les seules entreprises gagnant de l'argent sont les organisateurs de conférences et les éditeurs de magazines)*
- 3. la désillusion (la technologie ne va pas jusqu'au bout des attentes, elle devient vite démodée ; l'intérêt des media s'effrite)*
- 4. le chemin de la clarification (la réduction du champs d'expérimentation et le travail acharné réalisé par certaines sociétés amènent des applications technologiques ainsi que des risques et bénéfices associés ; des solutions logicielles industrielles accélèrent le déploiement des projets tout en en réduisant les coûts)*
- 5. le plateau de productivité (les réels bénéfices de la technologie sont démontrés et acceptés ; les outils et méthodologies sont de plus en plus stables ; la taille finale du plateau varie selon que la technologie est largement applicable ou qu'elle bénéficie simplement d'un marché de niche ; à ce stade on estime que 30% de l'audience cible à adopté la technologie une fois arrivée sur ce plateau) »¹⁹*

¹⁹ http://www.wmaker.net/lamachinecafe/Hype-cycle-of-emerging-technology_a119.html



En 2006, le web sémantique apparaît sur leur « cycle de maturité des technologies émergentes » :

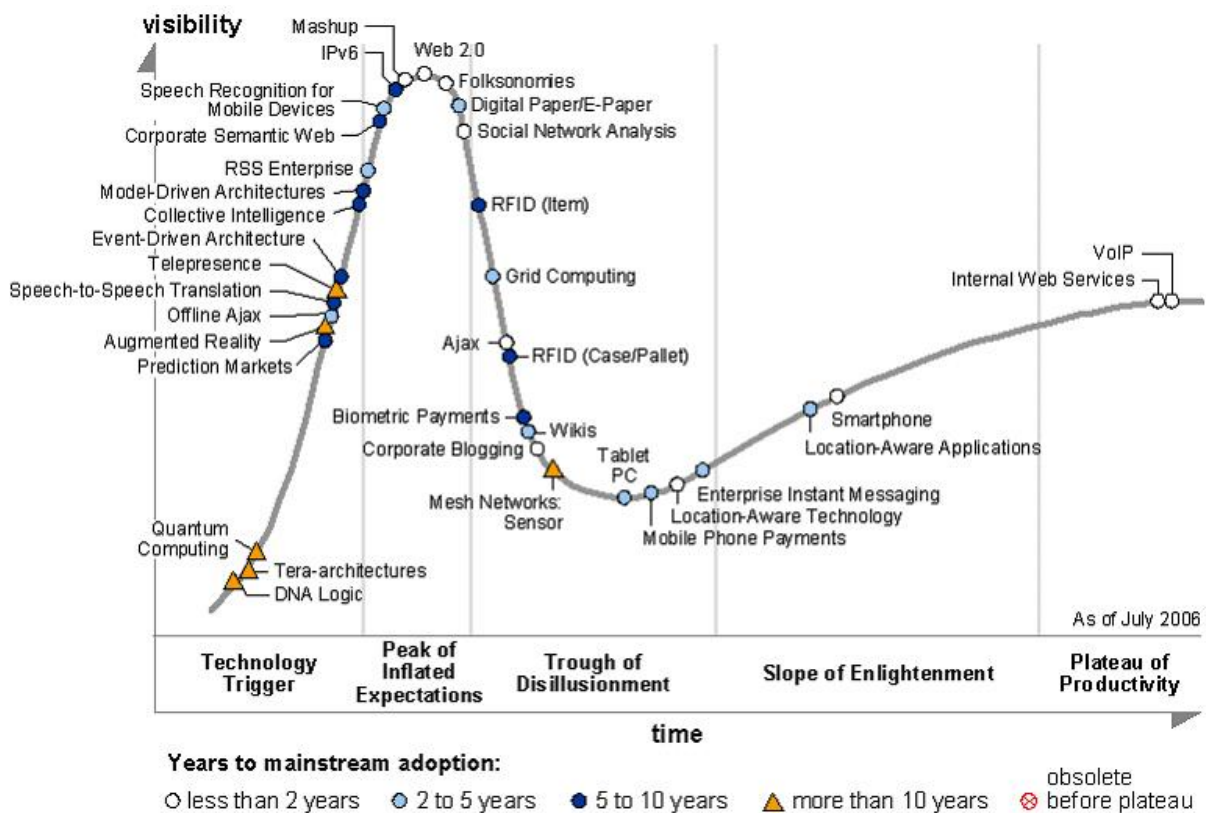


Figure 12 : Courbe du Gartner Group

Il est sorti de la phase de démarrage de la technologie pour entrer dans celle des expérances exagérées. Gartner Group prévoit alors son apogée d'ici 5 à 10 ans. On peut remarquer que le web 2.0 est en plein milieu de la phase d'enthousiasme et qu'il est attendu d'ici moins de 2 ans pour être une des principales technologies utilisées.

Hype Cycle for Emerging Technologies, 2007

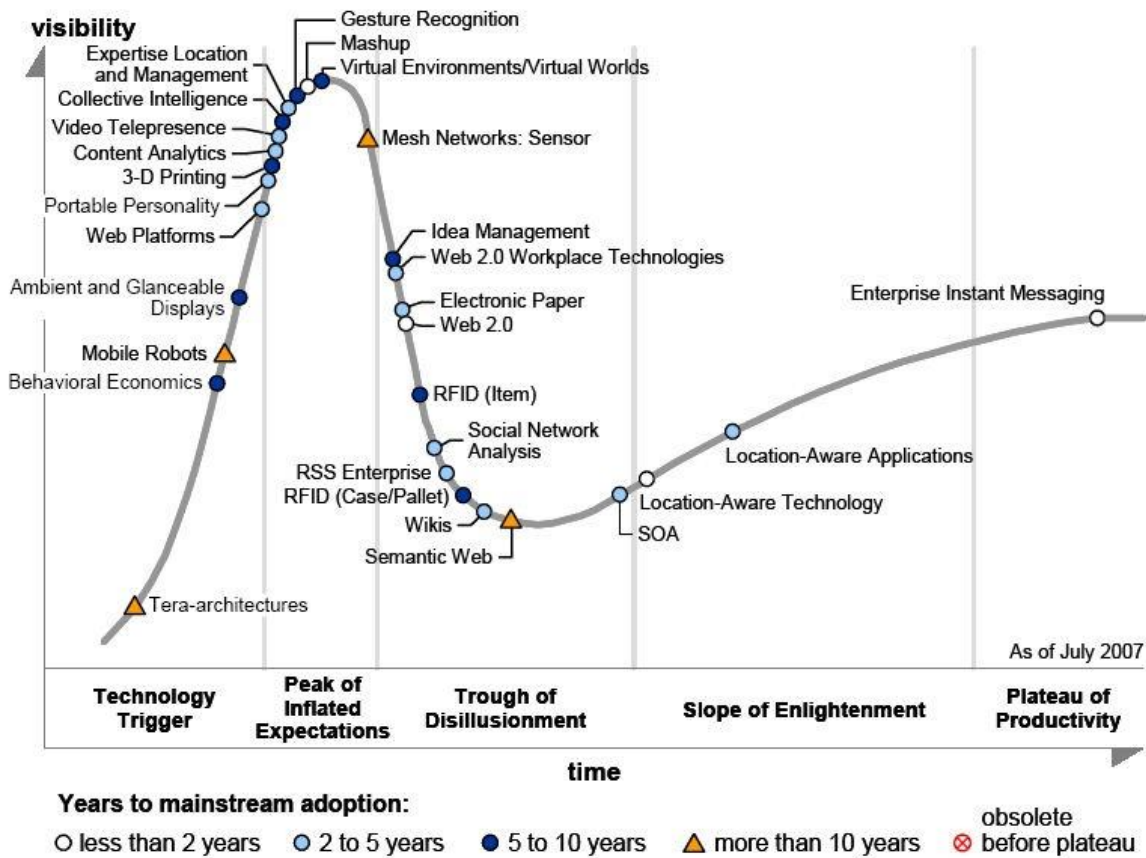


Figure 13 : cycle de maturité des technologies émergentes 2007- Gartner Group

En 2007 Gartner Group est plus pessimiste à propos du web sémantique. Celui-ci traverse un période de désillusion. Cette technologie n'est plus attendue avant au moins 10 ans. Le web 2.0 se trouve également dans cette phase, mais il est toujours attendu dans les 2 ans à venir.

Figure 1. Hype Cycle for Emerging Technologies, 2008

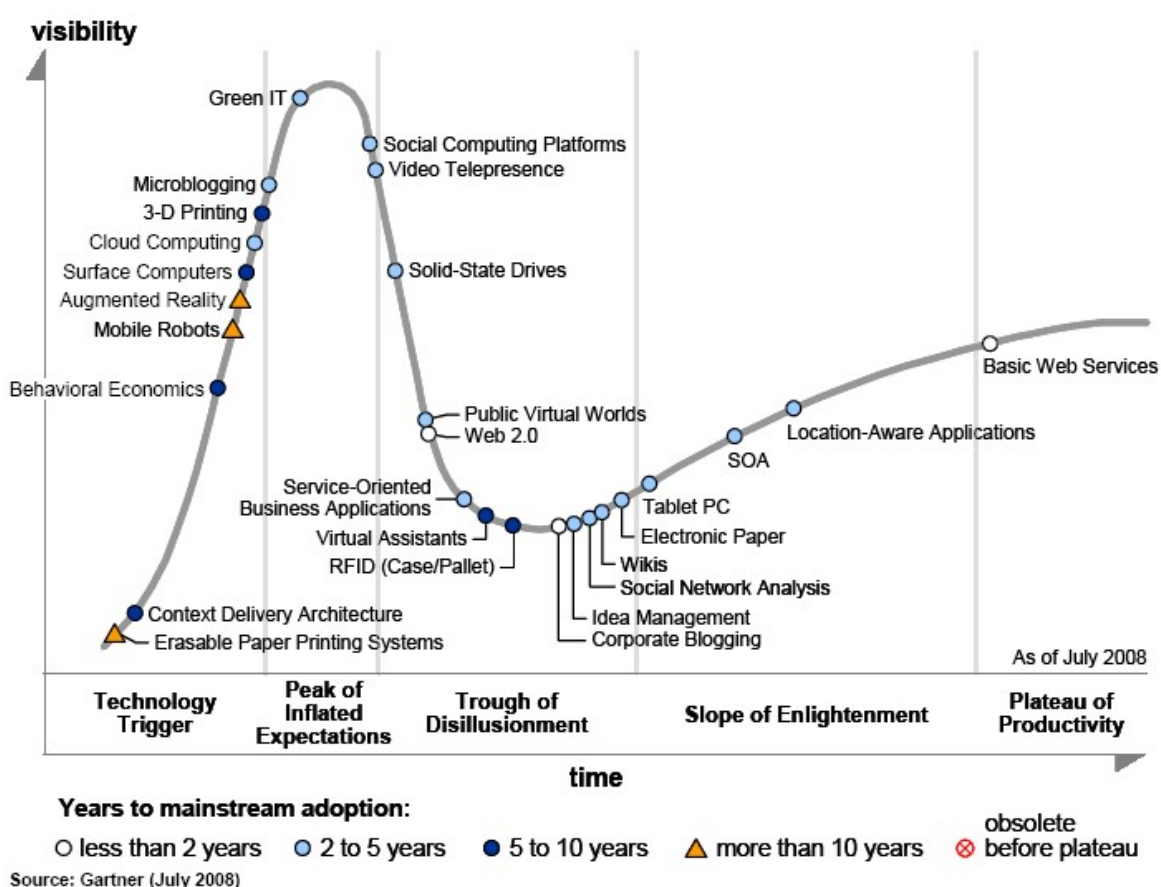


Figure 14 : cycle de maturité des technologies émergentes 2008- Gartner Group

En 2008, le web sémantique n'apparaît plus sur la courbe mais Gartner Group définit, la même année, la gestion des métadonnées comme faisant partie des 10 technologies qui vont marquer les tendances dans le business et les TIC.

En 2009, le web sémantique ne réapparaît toujours pas sur la courbe du Gartner Group.

Le web sémantique n'est pour l'instant connu que des spécialistes ou des personnes s'intéressant de près aux évolutions du web et de l'informatique en général. Il est pour l'instant inconnu ou presque du grand public. Dans certains domaines cependant (essentiellement scientifiques comme les sciences du vivant, la médecine, les mathématiques ou encore la linguistique), les outils sémantiques sont très utilisés et deviennent indispensables. Toutefois, l'arrivée du html 5 (dont on peut déjà apprécier les prémices) va rendre le web un peu plus sémantique. En effet le html 5 est basé sur l'utilisation de balises avec une notion sémantique qui permettra aux moteurs de recherche de mieux identifier les contenus. Par exemple il existe des balises <video> pour intégrer des vidéos, des balises <footer>, <article>,etc. Peut être est-ce un début vers un web plus sémantique et une ouverture vers web sémantique au grand public ?

On pourrait se demander l'utilité des ontologies quand on dispose de langage tel que UML pour modéliser les données et les traitements. Pourtant les deux langages sont complémentaires. L'UML s'intéresse aux applications, les données décrites et concervées dans les bases de données alors que l'ontologie est peuplé, c'est-à-dire que le contenu est déjà dans l'ontologie. Il est donc intéressant d'intégrer les deux schémas lors de la conception d'un programme.

4. Méthodologies et outils

Il existe aussi plusieurs outils pour développer et maintenir des ontologies. Voici certains de ces outils qui sont largement utilisés par la communauté des ontologistes.

- Ontolingua

Développé à l'*Université de Stanford*, le serveur Ontolingua est le plus connu des environnements de construction d'ontologies car il consiste en un ensemble d'outils et de services qui supportent la construction en coopération d'ontologies et particulièrement entre des groupes dont la situation géographique est différente.

- Ontosaurus

Développé à l'*Information Science Institut de l'Université de Southern California*. Ontosaurus est un serveur qui utilise le langage de représentation de connaissances Loom, ainsi qu'un serveur de navigation qui permet de créer dynamiquement des pages html.

- ODE

Développé au *laboratoire d'Intelligence Artificielle de l'Université de Madrid*, ODE (*Ontology Design Environment*) présente comme avantage principal d'offrir un module de conceptualisation pour construire des ontologies et un autre module pour construire des modèles.

- OILEd

Editeur d'ontologie graphique développé par l'Université de Manchester, il permet aux utilisateurs de développer des ontologies dans le langage DAML+OIL. Un aspect principal de l'outil OILEd est l'utilisation du raisonneur FaCT pour classifier des ontologies et vérifier leur consistance par la traduction de DAML+OIL à la logique de description en SHIQ.

- OntoEdit

Projet développé à l'Université de Karlsruhe, OntoEdit fournit des méthodologies et des processus d'intégration d'ontologie. Son éditeur d'ontologie permet à l'utilisateur de modifier la hiérarchie de concepts, ou encore d'implémenter des interfaces graphiques facilitant le processus de construction de l'ontologie (ex : la réorganisation de la hiérarchie)

- Protégé

Outil d'édition d'ontologie développé à l'Université de Stanford, il est utilisé largement aujourd'hui pour élaborer des ontologies en RDF(S) et OWL. L'outil fournit un environnement de développement graphique et interactif pour aider les ingénieurs et les experts du domaine à réaliser des tâches plus rapidement. Un des avantages de Protégé est son architecture ouverte et modulaire qui facilite le développement de fonctionnalités nouvelles à travers des *plugins-in* pour effectuer des opérations différentes.

Un tutorial très bien fait permet de comprendre comment créer facilement une ontologie avec Protégé. Nous pensons inutile de recopier ici les différentes étapes pour avancer pas à pas dans la création de l'ontologie, c'est pourquoi nous mettons directement le lien menant au dictaiel : <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf> :

Création, description et définition de classes/sous classes, propriétés/sous propriétés (inversées, symétriques, fonctionnelles, transitives), relations entre individus, domaines et rangs des propriétés, Classes primitives, restrictions et classifications automatiques.

5. OWL vs OWL2

Comme défini sur le site du W3C, *"The OWL 2 Web Ontology Language is an ontology language for the Semantic Web [...] compatible with the OWL standard of 2004 (now referred to as "OWL 1") which it supersedes. Like in OWL 1, the main syntactic form of OWL 2 ontologies is based on an RDF serialization, although various alternative syntactic forms are available too."*

Les principales différences entre OWL2 et OWL résident dans la façon de construire les ontologies : Pour certaines personnes de la communauté scientifique, OWL est un formalisme beaucoup trop complexe et détaillé, alors que pour d'autres, il est insuffisant et assez limité : OWL2 propose donc une nouveauté, un système de "profils" qui va permettre de n'utiliser qu'un sous-ensemble de OWL tout en restant interopérable.

Néanmoins, selon Ian Horrocks, professeur à l'université d'Oxford et animateur du Groupe de travail OWL, *"la nouvelle version d'OWL apporte puissance et rapidité, elle standardise les caractéristiques les plus demandées par les utilisateurs et présente des profils pour améliorer l'extensibilité dans les applications typiques."* Il s'agit ici de préciser que OWL vise à permettre la capture de connaissances sur un domaine spécifique afin de gérer, exploiter ces informations et faire des recherches. OWL est un outil qui doit permettre de créer et de partager ces vocabulaires. Au fil du temps, les connaissances évoluent. Il est également courant d'envisager de nouveaux modes d'utilisation des données, ou de combiner ces informations à d'autres par des biais non envisagés de prime abord. *"Ces réalités ont été prises en compte lors du développement d'OWL"* selon le W3C. Enfin, OWL permet de réduire les coûts de développement et de faciliter la conception de logiciels. Par exemple, OWL présente une fonction pour déduire que deux éléments d'intérêt sont identiques : *"la « planète Vénus » est également « l'étoile du matin » et « l'étoile du soir ».* Sachant que ces deux éléments représentent « une seule et même chose », les outils intelligents peuvent inférer des relations automatiquement, sans aucune modification du logiciel" selon le W3C, qui a publié la première version d'OWL en 2004. OWL est déjà utilisé dans des applications éparpillées telles que le commerce électronique, l'exploration du pétrole, la gestion des dossiers médicaux, les interfaces sémantiques ou la gestion d'archives musicales.

Les nouvelles fonctionnalités dans OWL 2 ont été apportées suite à une demande générale des utilisateurs d'OWL 1 : OWL 2 présente les profils OWL, sous-ensembles de langage facilitant l'utilisation et l'implémentation (aux dépens de l'expressivité) conçus pour répondre à différents besoins applicatifs, c'est ici que réside sa principale valeur ajoutée par rapport à OWL1.

Webographie

<http://public.web.cern.ch/public/fr/About/WebStory-fr.html>
<http://www.linternaute.com/histoire/motcle/3654/a/1/1/web.shtml>
http://fr.wikipedia.org/wiki/Histoire_d%27Internet
<http://naofan-creations.blogspot.com/2009/01/histoire-du-web-x-sommaire.html>
<http://www.commentcamarche.net/contents/histoire/ordinateur.php3>
<http://www.camilleroux.com/cv/>
<http://www.journaldunet.com/solutions/chat/retrans/070326-nitot-mozilla-europe.shtml>
<http://www.scientificamerican.com/article.cfm?id=the-semantic-web>
<http://www.cyberie.qc.ca/chronik/20010508.html>
<http://www.republique-des-lettres.fr/1230-tim-berners-lee.php>
<http://www.memoireonline.com/05/08/1145/conception-ontologie-plate-forme-enseignement-a-distance.html>
<http://www.inrialpes.fr/exmo/cooperation/asws/ASWS-Langages.pdf>
http://lacot.org/public/introduction_a_owl.pdf
<http://websemantique.org/PagePrincipale>
<http://tel.archives-ouvertes.fr/docs/00/19/87/18/PDF/These-final-Luong.pdf>
<http://www.webmaster-hub.com/publication/Applications-des-outils.html>
<http://www.paperblog.fr/2325639/comment-fonctionne-un-moteur-de-recherche/>
http://www.grappa.univ-lille3.fr/polys/info_ling/linguistique005.html
<http://www.webmaster-hub.com/publication/Applications-des-outils.html>
<http://wikiwix.com/cache/?url=http://wiki.crao.net/index.php/SeptIllusionsDeLaM%25E9ta>
<http://www.elanceur.org/Articles/OntologieSurfaite.html>
<http://eranksolution.net/blog/2009/05/wolfram-alpha-moteur-en-langage-naturel/>
<http://www.tunibox.com/divers/wolfram-alpha-le-premier-moteur-semantique-qui-revolutionne-le-web.html>
<http://www.journaldunet.com/solutions/moteur-referencement/selection/5-moteurs-de-recherche-en-langage-naturel/5-moteurs-de-recherche-en-langage-naturel.shtml>
<http://www.journaldunet.com/solutions/breve/international/39126/wolfram-alpha---naissance-d-un-moteur-de-recherche-encyclopedique.shtml>
<http://www.zdnet.fr/actualites/internet/0,39020774,39502148,00.htm>
http://www-sop.inria.fr/acacia/afia2005/actes/ActesWSeL_Final.pdf
<http://www.isetso.rnu.tn/cemaforad3/doc/Sandid%20Hattem.pdf>
<http://www.automatesintelligents.com/echanges/2008/jui/dossieria.html>
<http://www.web-libre.org/dossiers/intelligence-artificielle,7795.html>
<http://www.maxime-viry.com/blog/intelligence-artificielle-evolution-revolution>
http://doelan-gw.univ-brest.fr:8080/~dmasse/cours/cours_intro.pdf
http://fr.wikipedia.org/wiki/Syst%C3%A8me_expert
http://agora.unige.ch/tecfa/publicat/schneider/these-daniel/wmwork/www/phd_55.html
http://semantic-mediawiki.org/wiki/Help:Introduction_%C3%A0_Semantic_MediaWiki
<http://eranksolution.net/blog/2009/03/powerset-moteur-de-recherche-semantique/>
<http://tech.blogs.challenges.fr/archive/2008/07/01/avec-powerset-microsoft-met-un-pied-chez-wikipedia.html>
<http://www.gartner.com/it/page.jsp?id=530109>

<http://www.developpez.net/forums/d828672/club-professionnels-informatique/actualites/owl-2-nouveau-standard-w3c-web-semantique/>

http://semanticweb.org/wiki/OWL_2

<http://www.itrmanager.com/articles/96760/owl-2-standard-w3c-faciliter-integration-gestion-informations.html>

<http://protege.stanford.edu/overview/>

Glossaire

Linguistique : Science qui a pour objet l'étude du langage et des langues²⁰

Sémantique :

- Étude scientifique du langage du point de vue du sens
- Qui concerne la relation des signes et de leurs référents

Polysémie : qualité d'un mot qui présente plusieurs sens

Aide a la décision : Conception d'outils informatiques et définition de méthodes dans le but d'aider les décideurs à analyser un problème et à déterminer une solution.

Data mining : Exploration ou fouille de données. Procédure qui consiste à extraire et analyser des données inconnues.

Taxonomie : Science qui décrit des êtres vivants pour les identifier, les nommer et les regrouper en entités (Taxons).

²⁰ Dictionnaire Larousse