REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU

FACULTE DE GENIE ELECTRIQUE ET D'INFORMATIQUE

DEPARTEMENT D'INFORMATIQUE

MÉMOIRE DE FIN D'ÉTUDE

En vue de l'obtention du diplôme de Master en Informatique

Spécialité : Conduite de Projets Informatique

Thème : Eléments du Web et état de l'art

Réalisé par :

Fayçal Rèdha SAIDANI

Proposé et dirigé par :

M^r^ Idir RASSOUL

Devant le Jury composé de :

M^r^ Mohammed OUAMRANE	**Président**
M^r^ Aoamar OULARBI	**Examinateur**
M^r^ Mohamed RAMDANE	**Examinateur**

Promotion: 2011

Á mes Parents.

REMERCIEMENTS

Je tiens tout d’abord à exprimer ma gratitude à mon promoteur M^{r} Idir Rassoul, pour la confiance qu'il m'a accordé en me proposant ce sujet. Je le remercie également pour ses conseils, sa disponibilité mais aussi pour ses critiques toujours constructives et qui m’ont été d'un grand apport pour la finalisation de ce travail.

Je tiens également à remercier les membres de mon jury. Un grand merci à M^{r} Mohammed Ouamrane pour avoir accepté de présider le jury de ma soutenance, ainsi qu’à M^{r} Oularbi et Ramdane pour m’avoir fait l’honneur d’êtres examinateurs de ce mémoire.

Ma sympathie et mes meilleurs sentiments d'amitié vont vers tous les gens qui de près ou de loin m’ont soutenu durant la préparation de ce travail. Je pense en particulier à mon frère, mes sœurs et mes amis, notamment Khaled et Nassim.

Enfin, ces remerciements ne seraient pas complets sans l'expression de mon très profond amour pour mon père et ma mère, pour l'ensemble de ce que je leur dois, et leur affection. Merci infiniment.

TABLE DES MATIÈRES

INTRODUCTION GÉNÉRALE

Depuis son avènement au début des années 1990, le Web a révolutionné le monde de l'information, principalement en offrant un accès universel à la connaissance. Cette popularité a rapidement fait de cet outil la plus vaste base de données existante, de part la quantité et la diversité des documents qu'elle contient. Aussi, ces dernières années ont vu la montée en puissance de plusieurs visions du Web. Une des celles que nous vivons aujourd'hui, communément appelée par, **Web Participatif**, met l'accent sur la place centrale de l'utilisateur au sein des services Web ; elle met en avant les échanges, l'ouverture et la collaboration entre internautes par l'intermédiaire d'outils tels que les wikis, les blogs, les réseaux sociaux, les techniques de syndication etc. Toutefois, ces pratiques participatives nous confrontent inévitablement à un certain nombre de problèmes, notamment celui de la surcharge d'informations et qui, par conséquent, se répercutent négativement sur la qualité et l'intégrité de l'information disponible sur le Web. Une autre vision, celle du **Web Sémantique**, vise à structurer cette importante quantité d'informations, de manière à la rendre compréhensible par les machines et permettre ainsi des recherches plus précises et une navigation plus efficace ; Ceci en définissant des formalismes de représentations unifiées pour les données, dans un but d'échange et de compréhension de celles-ci par les agents logiciels.

D'autres avancées enregistrées dans les domaines de l'informatique mobile et de l'intelligence ambiante, nous permettent de voir progressivement apparaître une toute nouvelle approche, celle du **Web Ubiquitaire**. Ce terme symbolise une omniprésence du Web dans le monde réel, cette omniprésence permet aux internautes d'avoir accès à un ensemble de services au travers d'interfaces se voulant intelligentes. Ces dernières s'appuient principalement sur des technologies telles que les puces RFID[1] les codes à barres où encore les balises infrarouges, qui seront, voir très bientôt, intégrées et enfouîts dans chaque objet manufacturé. Elles supposent aussi une forte prise en compte du contexte dans lequel les utilisateurs évoluent (le lieu, la position sociale ou hiérarchique) pour adapter ces services Web en fonction de leurs comportements, et prendre en compte dynamiquement l'arrivée de nouveaux éléments dans l'environnement (utilisateurs ou dispositifs). Á ce propos, de récents travaux de recherches ont vus le jour [Tandabany, 2009], [Schmidt, 2010]. Cependant, le Web actuel est encore loin de cet idéal.

1. Méthode utilisée pour stocker et récupérer des données à distance en utilisant des balises métalliques.

Il existe toutefois, plusieurs autres termes permettant de décrire ces différentes visions. Parmi les plus connus et les mieux acceptés au sein des communautés Web, on retrouve les termes « Web 1.0 », « Web 2.0 », « Web 3.0 » et voir même « Web 4.0 ». Mais contrairement à ce que leurs suffixes laissent croire, ces termes ne font pas référence à des versions, mais bien à la reconnaissance d'une étape nouvelle du Web. Le Web 1.0 fait par exemple référence à cette phase où les pages étaient principalement statiques, et où l'internaute n'était qu'un simple consommateur d'informations. Le Web 2.0, quant à lui, fait référence aux pages dynamiques, où l'internaute contribue à l'échange d'informations et interaction, à la fois avec le contenu et la structure des pages.

La suivante étape évolutive du Web, soit le « Web 3.0 », commence peu à peu à se développer. Elle fait référence à un Web fortement structuré et où chaque donnée est assignée à une autre métadonnée appelé Annotation Sémantique. Cette étape, où du moins l'un des axes de son développement, est appelé Web Sémantique ou encore Web de données, et représente l'un des projets phares du W3C. Enfin pour résumer, le « Web 3.0 », « Web sémantique » ou « Web des données » peut se voir comme un Web à travers lequel les technologies parviennent à mieux répondre aux besoins et aux usages des internautes, et celà, sans que ces derniers ne s'en aperçoivent.

Aussi, l'objectif de ce mémoire est de situer les différents concepts qui caractérisent chaque version du Web. Nous nous attacherons à mettre en évidence les faiblesses dont souffre actuellement chacune d'elles, tout en relatant les divers travaux de recherches sur le sujet.

Ce document est organisé en quatre chapitres.

CHAPITRE 1 : HISTORIQUE SUR LES CONCEPTS ÉMERGENTS DU WEB

Dans ce chapitre, nous présentons sous forme chronologique un aperçu de chacune des quatre versions Web. Nous aborderons aussi les trois différents paradigmes que l'on rencontre régulièrement.

CHAPITRE 2 : ETAT DE L'ART DU WEB 2.0

Dans celui-ci, nous essayerons d'éclaircir le concept Web 2.0, en décrivant ses sept piliers fondateurs énoncés dans le document référence[2] de Tim O'Reilly[3], nous poursuivrons par une énumération des principaux usages et techniques ayant émergé durant cette période de

1. http://oreilly.com/web2/archive/what-is-web-20.html
2. fondateur d'O'Reilly Media, et premier à avoir utilisé le terme Web 2.0

Web 2.0. Nous discuterons aussi sur les principales limites de ce dernier. Nous terminerons enfin par présenter le « Web Squared » qui correspond selon Tim O'Reilly au Web de demain.

CHAPITRE 3 : VERS L'ÈRE DU WEB 3.0

Ce chapitre se divise en trois sections ; la première concerne les enjeux techniques du Web sémantique ; la seconde, traite du projet OPEN DATA, qui représente une approche complémentaire au Web sémantique. Et la troisième, sera un bref aperçu sur le *Cloud Computing*, pour lequel il sera seulement question de prendre conscience des différentes possibilités qu'il pourrait nous offrir.

CHAPITRE 4 : LE WEB 4.0 EN PERSPECTIVE

Dans ce chapitre, nous laissons entrevoir en guise de perspective, certains aspects qui pourraient s'avérer êtres la clé maitresse pour voir un jour émerger un Web d'objets communicants, tel qu'il est décrit par Joël de Rosnay[3].

3. Spécialiste des origines du vivant et des nouvelles technologies, puis en systémique et en futurologie (ou prospective).

CHAPITRE 1 : HISTORIQUE SUR LES CONCEPTS ÉMERGENTS DU WEB

Ce chapitre expose succinctement les différentes phases marquant l'évolution du Web, ainsi que les différents paradigmes qui leurs sont attachées.

1.1. Chronologie de l'évolution Web

Les dénominations Web 1.0, Web 2.0 ou encore Web 3.0 ne font pas référence à un numéro de version, contrairement à ce que l'on pourrait penser. Le Web est multiple et c'est donc très librement que nous désignons aujourd'hui par Web 1.0 les premiers usages d'Internet, comme pour marquer l'antériorité de l'un par rapport à l'autre. Il existe encore bien d'autres appellations, chacune rendant compte de particularités remarquables. Parmi ces dénominations, nous en avons retenu trois, celles qui se rencontrent régulièrement à savoir, le Web Statique, le Web Participatif et le Web Sémantique. La figure 1.1. Illustre clairement ces différentes phases :

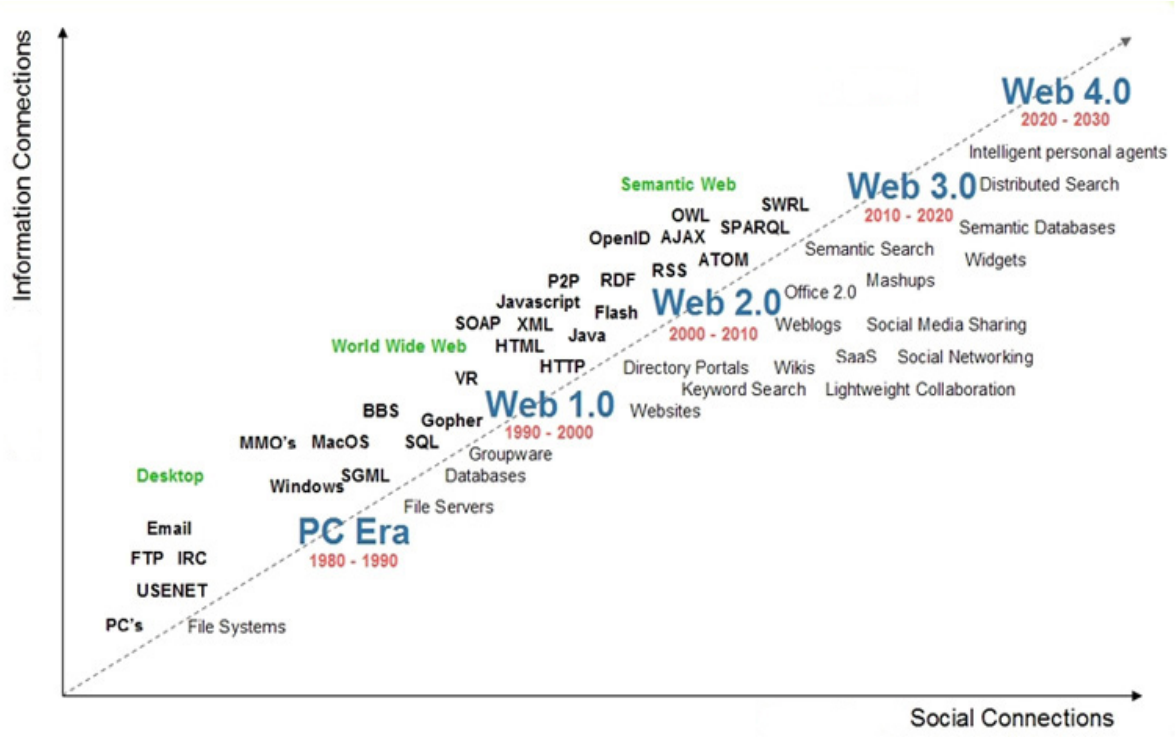

Figure 1.1. Les grandes étapes de l'évolution du Web[4].

4. Illustration de Nova Spivack. (Source : www.radarnetworks.com)

1.1.1. Web 1.0

Nous qualifions par Web 1.0, la période où il était rendu possible, aux internautes les plus avertis, de publier des pages HTML (*HyperText Mark-Up Language*) mélangeant, du texte, des liens, des images, le tout consultable en ligne dans un navigateur Web grâce au protocole HTTP (*Hypertext Transfer Protocol*). On estime que cette période s'étend de 1994 jusqu'à 1997.

Toutefois, ces sites n'étaient pas différents les uns des autres, et se contentaient juste de présenter des informations d'une façon plus ou moins structurée. Ce web est d'ailleurs souvent appelé « *Web des consommateurs* » vu l'infime nombre d'interactions qui pouvait exister entre un site et son visiteur.

Le Web 1.0 est ainsi caractérisé par :

- La notion de « Site Web Vitrine » qui est comparable dans le monde physique à une bibliothèque où il faut se rendre pour avoir accès à son contenu et dans laquelle un individu ne peut pas modifier une information mais uniquement la consulter ;

- La possibilité de publier un contenu Web uniquement par le propriétaire du site Web ;

- L'attitude passive de l'internaute qui ne peut que consulter les pages.

1.1.2. Web 1.5

Au cours des années 1997 → 2001, le Web a évolué au niveau 1.5, quelque peu dynamique. Il devient alors possible par exemple de consulter des sites recensant des milliers d'informations sous la forme d'un catalogue régulièrement mis à jours, c'est à cette période là aussi, qu'ont vu le jour les premiers sites de commerce en ligne. Ce Web dynamique est généralement basé sur l'association du langage de programmation PHP *(Hypertext PreProcessor)* et des bases de données Mysql. Lorsque l'internaute accède au site dynamisé, il fait exécuter sur le serveur le langage PHP qui va chercher l'information dans la base de données pour la retranscrire dans la page HTML. Les forums et chats font partie des seuls outils interactifs qui ont vu le jour à cette période.

1.1.3. Web 2.0

Quelques temps après les années 2001, le Web a subi d'énorme transformation, notamment après l'explosion de la bulle Internet en 2000 et la prise de conscience du potentiel de cette technologie. Les services se sont alors développés, devenant de plus en plus nombreux, mais surtout de plus en plus interactifs. Ainsi, l'Internaute n'est plus simplement consommateur d'information, mais peut s'il le souhaite devenir producteur. Et tout est mis en œuvre pour qu'il le devienne. Cette révolution a donné naissance à ce que l'on nomme aujourd'hui le Web 2.0 appelé aussi Web Social, Web Participatif ou encore Web collaboratif.

Une des caractéristiques principales du Web 2.0 qui le distingue grandement du Web 1.0 est la prise de contrôle de l'information par les utilisateurs. N'importe quel internaute peut aujourd'hui se faire une place sur la toile, il peut collaborer, partager des informations, des outils, des fichiers multimédias, donner ses opinions, commenter, réagir etc, tout ceci sans connaissances spécifiques en informatique. En effet, quand auparavant il fallait un minimum de savoir faire en programmation pour créer son espace sur le Web, aujourd'hui cela n'est plus nécessaire vu le nombre d'outils qui sont mis à disposition de tout un chacun afin de faciliter toutes ces interactions.

Parmi ces outils, nous pouvons bien entendu citer les réseaux sociaux, les blogs, les wikis, les sites de partage de vidéos, de photos, de musiques, etc. La grande majorité des sites présents sur le Web aujourd'hui offrent la possibilité à tous leurs visiteurs de laisser, au minimum, une trace textuelle de leurs passages sur le Web. Tout ce contenu, qu'il soit textuel ou autre, est appelé *User Generated Content*.

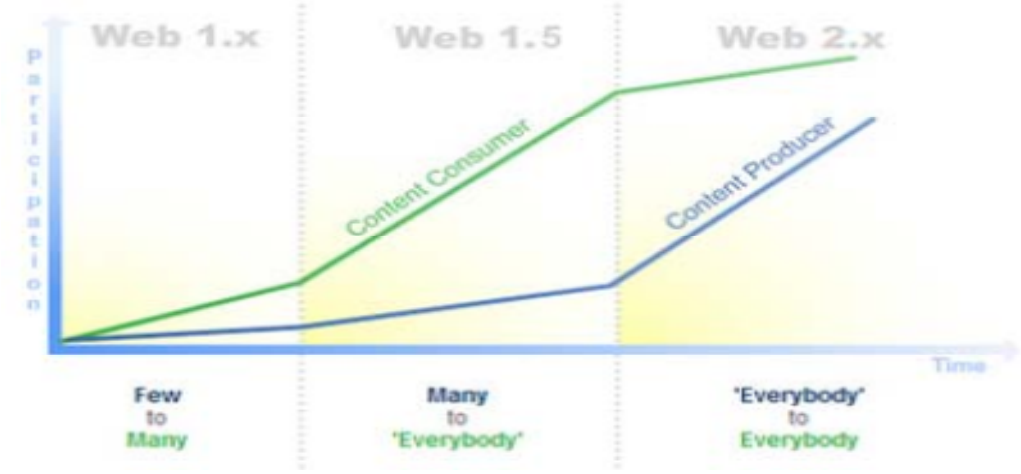

Figure 1.2. Analogie des usages Web avant le Web 2.0[5]

1.1.4. Web 3.0

Il est difficile de parler de Web 3.0, car des appellations concurrentes sont d'ores et déjà utilisées, telles que le « Web sémantique », d'autres appellations seraient aussi tout à fait valables, par exemple « Web squared » qui aurait l'avantage de marquer à la fois la continuité avec le Web 2.0 et l'accroissement de la puissance et des capacités de ce Web futur. Néanmoins, quel que soit le nom qui sera adopté, le Web 3.0 reste une étape dans laquelle, toutes les communautés Web devrons faire face à la problématique de l'explosion quantitative de données engendrée par le Web 2.0, et ainsi mettre en place des solutions qui se doivent d'être en conformité avec le W3C[6].

Á ce propos, les théories et les technologies du Web sémantique sont aujourd'hui presque toutes standardisées et sont en phase d'être exploitées. Le Web sémantique a pour

5. www.bitsandbuzz.com
6. The World Wide Web Consortium, www.w3.org

rôle de construire des outils d'assistance aux utilisateurs, qui automatisent le décryptage du sens des données non structurées (textes en langues naturelles, images, vidéos) pour apporter des réponses pertinentes aux requêtes. Cela tout en respectant le point de vue des utilisateurs.

L'expression Web 3.0 pourrait bien aussi faire référence à une autre évolution, qui cette fois-ci serait omniprésent et indépendant de tout type de support matériel pour fonctionner, ce concept pourrait bien se mettre en place grâce aux avancées enregistrées dans les domaine du *Cloud Computing*. Ainsi les bases du Web 3.0 se résumeraient à un Web aux performances de recherches améliorées, à une portabilité et une mobilité sans précédent d'internet. La figure ci-dessous schématise ces principales caractéristique du Web 3.0 :

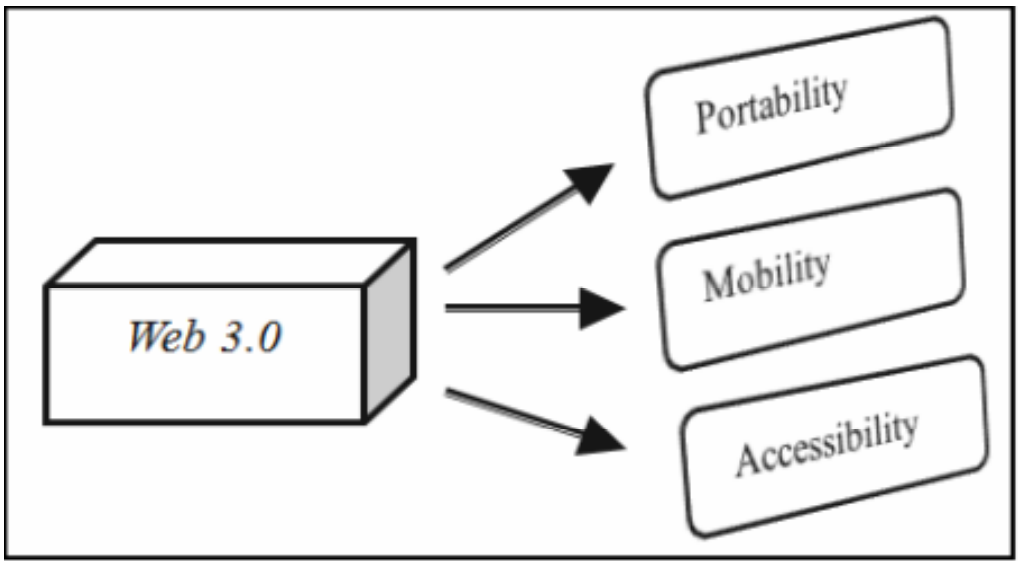

Figure 1.3. : Les bases du Web 3.0.

1.1.5. Web 4.0

Il fait référence à ce que l'on appelle le Web des Objets. Ainsi le Web 4.0 permettra de connecter les gens et les objets n'importe où, n'importe quand, par n'importe qui, tous ceci grâce à des capteurs reliant l'univers physique à celui du numérique.

1.2. Les Différents Paradigmes du Web

Web statique, collaboratif, participatif, mobile, sémantique, ubiquitaire… ce sont autant de visions du Web, toutes différentes les unes des autres et qui contribuent toutes à son évolution, pour offrir ainsi à ses utilisateurs de plus en plus de fonctionnalités. Toutefois, nous nous intéresserons dans cette section à trois d'entre elles seulement, à savoir le Web participatif, l'ubiquitaire et enfin le Web sémantique.

1.2.1. Le Web Participatif

« *Le Web Participatif* » représente une phase du Web marqué par la participation croissante des internautes dans la production et la gestion de contenus. Ce terme recouvre en grande partie l'expression « Web 2.0 ». Mais contrairement au Web Sémantique que nous verrons par la suite et qui est lié à un changement technologique, le Web Participatif est principalement une évolution des usages liée à la démocratisation du Web. Cette participation s'illustre par des sites comme l'encyclopédie collaborative Wikipédia, les blogs, les wikis, les réseaux sociaux etc.

1.2.2. Le Web Sémantique

Présentation générale : Le Web sémantique est un ensemble de spécifications, langages et architectures proposés par le consortium international W3C. Ce consortium regroupe des membres d'universités, discutant et orientant les évolutions techniques du Web. Parmi ses membres, le plus connu est *Tim Berners-Lee*, l'inventeur du Web. En 2001, dans son article [Berners-Lee et al., 2001], Tim Berners-Lee et ses collaborateurs présentent le Web sémantique comme le futur du Web, Un Web dont le contenu est compréhensible par les machines.

Ainsi le Web sémantique, permet aux machines de comprendre la signification des informations présentes sur le Web. Il permet aussi d'étendre le réseau des hyperliens entre les pages Web classiques par un réseau de liens entre données structurées laissant ainsi la chance aux agents automatiques d'accéder plus intelligemment aux différentes sources de données Web, de manière à fournir des tâches (recherche, apprentissage, etc.) plus précises aux utilisateurs. Il englobe entre autre différents langages, parmi lesquels on retrouve *RDF, OWl, SKOS et SPARQL*. Quand à la sémantique des données, elle est décrite par des ontologies et certains des langages précédemment cités à savoir RDF et OWL.

Agenda de développement : Pour que le Web classique devienne Sémantique, le W3C a proposé un agenda sous la forme de composants techniques à développer : le fameux « Gâteau sémantique » (Fig. 1.4). Chaque composant se sert des précédents pour fonctionner. Les couches inférieures (URI, UNICODE, XML, RDF et RDF-S) ont été standardisées et sont largement employées sur le Web. Elles concernent des aspects syntaxiques, c'est-à-dire les formats à employer pour écrire, nommer et échanger des données de manière commune. Actuellement, les efforts de développement portent sur les couches médianes (SPARQL, OWL et RIF/SWRL). Les difficultés portent notamment sur les ontologies souvent coûteuses à concevoir et à rendre cohérentes entre elles.

Aussi, afin d'offrir une meilleur capacité de traitement automatique, il devra s'ajouter aux informations existantes sur le Web une couche de métadonnées, les rendant ainsi exploitables par la machine. Ces métadonnées apporteront une certaine sémantique pour automatiser les traitements voulus. Enfin, vu qu'il est important de saisir ces notions pour la

compréhension des chapitres à venir, nous proposant de voir plus en détail certains de ces éléments.

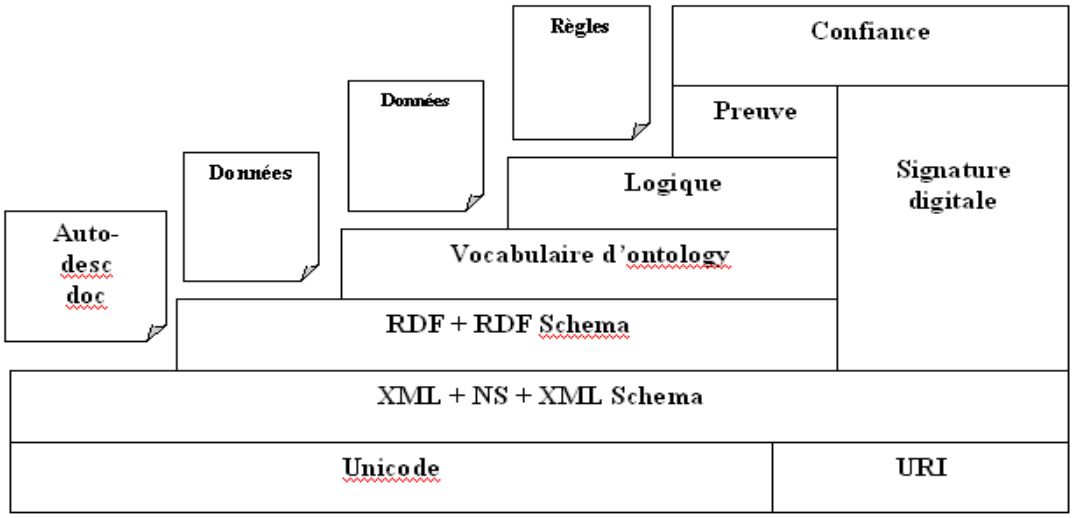

Figure 1.4. Figure illustrant le célèbre Semantic Cake.

- **Les Principales Composantes du Web Sémantique**

1.2.2.1. La couche preuve & confiance

Comme le Web est un espace d'information libre, la confiance d'une information trouvée n'est pas vraiment garantie dans tous les cas. Une telle confiance doit reposer sur des méthodes de qualification de l'origine de l'information ; soit à l'aide de métadonnées ou d'annotations sémantiques (Section 1.2.2.9, page 13) éventuellement certifiées par des signatures électroniques. De ce fait, les métadonnées RDF pourront être chiffrées dans le Web sémantique selon des techniques de signature connues.

De manière plus ambitieuse, la capacité de produire des preuves, pourra augmenter le niveau de confiance des utilisateurs dans ces déductions. Un langage de preuve est simplement ce qui nous permet de prouver si un rapport est vrai. Un exemple de langage de preuve se composera généralement d'une liste de faits qui ont été employés pour dériver l'information en question, et la confiance pour chacun de ces faits qui ont été vérifiés.

1.2.2.2. Couche logique & inférence

Afin de pouvoir résoudre toute variation de problèmes réels, le web sémantique devrait supporter un langage logique permettant des inférences plus ou moins complètes. En exploitant des connaissances disponibles sur le web sous forme de règles, les agents logiciels pourront ainsi raisonner intelligemment et offrir des réponses automatiques à des questions posées par une personne. Cette possibilité devra s'appuyer sur la standardisation d'un langage de règles adopté pour le Web sémantique, SWRL -*Semantic Web Rule Language*-, une combinaison d'OWL et RuleML *Rule Markup Language*.

1.2.2.3. Ontologies, Classes et Entités.

Une ontologie est une représentation de connaissances composée de classes et d'entités. Les classes sont des ensembles d'entités possédant des caractéristiques similaires. Une classe pour les machines est un peu l'équivalent d'un concept pour les humains.

Dans la Figure. 1.5, la première ligne pourrait être interprétée comme le nom de deux entités « Cours » et « Web sémantique » appartenant respectivement à des classes « Ressources pédagogiques » et « Disciplines informatique », appartenant elles-mêmes à une ontologie du domaine pédagogique.

Cours de Web sémantique	**mot1 mot2 mot3**
Par Tim Berner	**mot4 mot5 mot6**
Master	**mot7**
Les Ontologies	**mot8 mot10**

Figure 1.5. Extrait de contenu pédagogique, ce que lit un humain, à gauche, et ce que perçoit une machine, à droite.

La première caractéristique des ontologies est que les classes et les entités possèdent obligatoirement une dénomination pour les désigner (URI). Cette dénomination prend généralement la forme d'une adresse Web. Alors qu'en langage naturel, un mot comme « Master » peut désigner aussi bien une marque de véhicule qu'un niveau universitaire, dans le Web Sémantique, chacun de ces sens correspondrait à une classe désignable par une dénomination/adresse.
Exemple : « http ://UnConstructeurAutomobile.com/ontoVoiture#master » ou « http ://MinistereEducation.org/ontoEducation#master ».

Une seconde caractéristique des ontologies est que les classes et les entités peuvent posséder des relations entre elles et sur lesquelles les machines pourront effectuer des raisonnements automatisés. Par exemple, il pourrait exister une relation « enseigne » entre une classe « Humain » et une classe « Discipline informatique ». Cette relation permettrait à une machine de répondre à « quelle est la discipline enseignée par Tim Berner ? ». Une machine peut aussi inférer de nouvelles connaissances, c'est-à-dire déduire par exemple que si « Tim Berner » enseigne, alors celui-ci est un « Enseignant ». Par la suite, d'autres raisonnements pourront s'appuyer sur cette nouvelle connaissance et ainsi de suite.

- **Langage de Représentation de Connaissance**

En se basant toujours sur la Figure 1.4, On remarque que le Web sémantique s’appuie sur une pyramide de langages dont seulement les couches basses sont aujourd’hui relativement stabilisées. Il repose aussi sur la notion d’URI (Uniform Resource Identifier). Mais une des caractéristiques de tous ces langages est d’être systématiquement exprimable et échangeables dans une syntaxe XML.

1.2.2.4. Resource Description Framework – RDF –

RDF est un modèle de graphe destiné à la représentation et a une meilleure exploitation des métadonnées. Ainsi, de manière plus générale, « RDF permet de voir le web comme un ensemble de ressources reliées par les liens étiquetés sémantiquement »[6]. Le standard RDF représente donc un moyen d'écrire d'une manière standardisée, une assertion du type « ***r**essource-**a**ttribut-**v**aleur* ». Un tel triplet est interprétable comme une déclaration de trois entités « **s**ujet-**p**rédicat-**o**bjet ».

Précisément, le sujet en RDF « *ce sur quoi porte la déclaration* » est nécessairement un objet de type ressource (ex : page web) ou toute chose pouvant être référencée par une URI. Quant au prédicat, il doit être de type propriété. Il est lui-même identifié par URI. Chaque propriété possède une signification bien précise qui donnera la sémantique de description. Enfin, l'objet peut être une autre ressource mais aussi simplement une chaîne de caractères appelée littéral. Vu que RDF est un modèle de graphe, cela suppose qu'un document RDF ainsi formé correspond à un multi-graphe orienté étiqueté dont chaque triplet correspond alors à :

- Un arc orienté entre le label est le prédicat.
- Le nœud source est le sujet.
- Le nœud cible est l'objet.

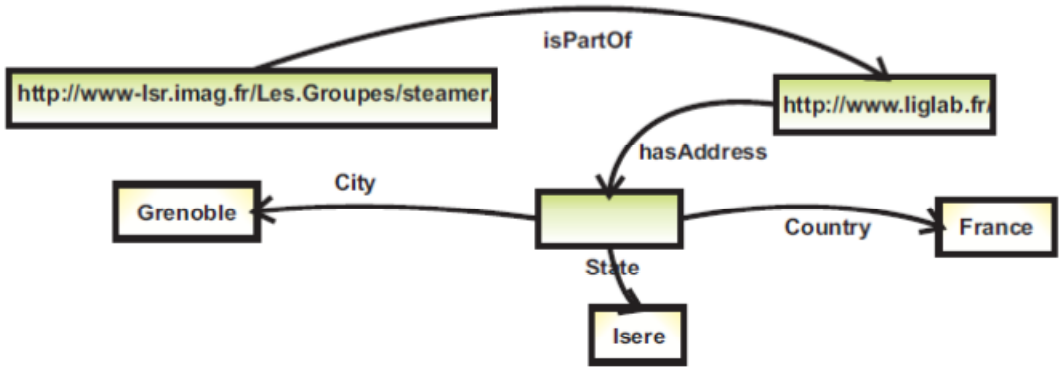

Figure 1.6. Exemple de graphe RDF [7].

Un des avantages de RDF vient de son extensibilité, à travers l'utilisation de schémas RDFs**.** C'est ce schéma qui donne véritablement sa sémantique à la description RDF, car il permet de décrire le sujet et prédicat dans un modèle objet. En effet, le schéma RDFS permet de définir de nombreux vocabulaires différents les uns des autres et pouvant être adaptés chacun à un domaine ou à une application spécifique. Notons que les vocabulaires, appelés aussi schémas de description, sont eux-mêmes écrits en RDF, en utilisant des balises de l'espace de nom RDFS.

7. www.w3.org

1.2.2.5. OWL - Ontology Web Language -

OWL est un standard du W3C, basé aussi sur RDFs. Il permet de définir les classes de manière plus complexe, c'est-à-dire que, contrairement à RDFs, il définit des classes avec des connecteurs de la logique de description et dispose donc d'une sémantique formelle claire, c'est ce qui lui permet de se doter de services inférentiels.

OWL est décomposé en trois sous langages : *OWL Lite, OWL DL et OWL Full*, qui sont conçus comme trois couches successives, c'est-à-dire en allant du plus simple (*OWL Lite*) au plus complexe (*OWL Full*).

- *OWL Full* correspond au langage OWL dans son ensemble, c'est-à-dire qui inclut tous les constructeurs d'OWL et toutes les possibilités de RDFs. L'inférence en OWL Full constitue ainsi un problème indécidable.

- *OWL DL* est un sous-langage d'OWL Full et ne permet pas l'utilisation de certaines constructions particulières de RDFs tel que la réflexivité etc.

- *OWL Lite* est un sous-langage d'OWL moins expressive. OWL Lite ne reprend de OWL DL que les constructeurs jugés les plus utiles et les plus faciles à mettre en œuvre.

1.2.2.6. SPARQL - Query Language for RDF -

SPARQL est un langage de requête proposé par le W3C et dédié à RDF, il est largement utilisé dans le domaine de recherche et d'extraction d'informations. Le langage SPARQL est basé sur la correspondance des patrons de graphe qui prennent une forme similaire aux triplets en RDF mais qui possède la capacité d'exprimer des variables de requête dans les positions du sujet, de la propriété ou de l'objet de ce triplet [W3C].

1.2.2.7. RDFa

RDFa est un langage dont l'objectif principal est de rajouter les informations en RDF dans les documents HTML ou XHTML ; celui-ci fournit une syntaxe et un ensemble de balises (tags) pour décrire des données structurées en XHTML, en plus de ces données structurées, sont rajoutées les sémantiques qui permettront l'échange d'informations par les applications automatiques ou par les agents informatiques.

1.2.2.8. RIF - Rule Interchange Format -

RIF est un formalisme permettant de fournir l'interopérabilité entre les langages de règles en général et ceux utilisés en particulier pour le Web. Le noyau de ce langage, *RIF Core*, correspond à la logique de Horn. Il offre certaines extensions inspirées des langages à objets et de frames ainsi que les URIs. La partie principale de *RIF Core* est le langage de

condition qui définit la syntaxe et la sémantique des règles de RIF ainsi que la syntaxe des requêtes. D'autre part, la spécification de RIF repose sur certains types de règles tels que les règles de production, de programmation logique et règles basées sur la logique du premier ordre.

1.2.2.9. Les Annotations Sémantiques

Une annotation (ou métadonnée) est une information descriptive facilitant l'accès, la recherche et l'utilisation d'une ressource. Elle se base sur un modèle de connaissances déjà défini (une ontologie) qui enrichit cette annotation en lui attribuant une sémantique et en la rendant utilisable comme telle par un agent logiciel. Ainsi, les annotations sémantiques décrivent le lien entre les entités se trouvant dans le document et leurs descriptions sémantiques représentées dans l'ontologie. Elles permettent ainsi de désambiguïser le contenu du document pour un traitement automatique.

1.2.3. Le Web Ubiquitaire

Le Web ubiquitaire est une autre perception d'internet, qui cherche à élargir les capacités des navigateurs Web afin de permettre de nouveaux types d'applications Web. Ces applications seront capables d'exploiter les services web pour élargir les capacités des terminaux. Les utilisateurs pourront alors se focaliser sur ce qu'ils font et non sur les terminaux eux-mêmes. La mobilité de ces applications permettra à toute personne de continuer à travailler tout en passant en douceur d'un terminal à l'autre.

L'intérêt du web ubiquitaire est qu'il rend facile la façon avec lequelle les développeurs créées leur applications, grâce notamment à l'utilisation combinée de balises, de graphiques, de feuilles de style et de scripts. Le Web ubiquitaire facilitera ainsi le développement d'applications distribuées en présentant des abstractions claires aux développeurs Web souhaitant accéder aux capacités des terminaux et aux services de communication. La découverte et la description des ressources seront essentielles à la création d'applications du Web ubiquitaire. L'utilisation des URI mais aussi les services et les sessions permettront l'emploi de métadonnées riches pour la découverte de ressources, intervenant à travers les différents réseaux et exploitant la nature distribuée du Web.

La différence essentielle entre; le Web actuel et celui Ubiquitaire ; est le besoin de mobilité, et par conséquent la prise en compte de l'environnement. En effet, une particularité du Web ubiquitaire est le besoin de communication d'un objet avec son environnement, ceci dans le but de pouvoir coopérer avec les objets qui l'entourent et ainsi accéder aisément à l'information. L'utilisateur sera alors pris en compte avec son contexte physique afin qu'il puisse avoir un accès mobile à des données et traitements Web, ce qui permettra de lui offrir les meilleures conditions de service. La mise en œuvre de ces services s'appuiera de façon prédominante sur des communications sans-fil. D'autre part, on notera que cette mobilité entraîne un besoin de reconfiguration dynamique des systèmes afin de s'adapter à tout instant

à l'environnement présent. Il s'agit donc d'intégrer des logiciels capables de s'auto-configurer aux objets utilisés quotidiennement.
Ainsi, nous distinguons deux principales particularités au Web Ubiquitaire :

- **La miniaturisation**

Le Web Ubiquitaire est notamment caractérisée par des objets sensibles à leurs environnement et capables d'interagir avec lui, afin de pouvoir communiquer à tout instant. Ainsi du matériel de plus en plus léger et spécialisé est donc nécessaires pour embarquer ces données de diverses natures.

- **L'interdépendance**

D'autre part, dans un tel Web, les interfaces Web classiques tels que nous les connaissons aujourd'hui devient quelque peu obsolète. En effet, car avec le Web ubiquitaire on étend la notion d'interface à tous les objets de l'environnement en vue de prendre en compte leurs usages intuitifs. Les exemples les plus simples sont l'arrêt de bus qui affiche les horaires et les itinéraires à suivre sur un PDA, ou encore une visite de musée qui permet au passage devant les différentes œuvres d'obtenir des informations détaillées sur ces dernières.

1.3. Conclusion

Après avoir subi d'énormes changements, tant du point de vue structurel que fonctionnel grâce aux pratiques participatives, Le Web enregistre aujourd'hui de plus en plus d'applications issues des pratiques du Web sémantique, lui permettant ainsi de renforcer les raisonnements automatisés sur les connaissances qu'il contient. Dans les prochains chapitres, nous verrons en détail chacune de ces évolutions, tout en précisant les technologies, services et offres caractérisant celles-ci, ainsi que les enjeux et problématiques de chacune d'elles.

CHAPITRE 2 : ETAT DE L'ART DU WEB 2.0

Après avoir longtemps été un simple espace de stockage à titre essentiellement consultatif, on assiste ces dernière années à une montée en puissante du Web participatif aussi appelée Web 2.0. Ce dernier modifie de manière profonde la façon dont les contenus sont publiés et échangés en ligne. Cette rupture concernant la production d'informations en ligne est notamment due aux outils dits participatifs.

Dans ce chapitre, nous présenterons le Web 2.0, tout en énumérant les outils phares de cette mouvance, à savoir les blogs, Wiki, les réseaux sociaux, les principes de syndication de contenu et enfin, la notion de tagging et les limites qu'elle comporte, ainsi que certains travaux de recherches et améliorations qui pourraient lui être apportées.

2.1. Les Sept principes du Web 2.0

Bien que sa définition n'a pas été clairement établie, de nombreux aspects caractérisant le Web 2.0 ont été mis en avant à l'issue de l'article paru en 2005 sous le nom « QU'EST CE QUE LE WEB 2.0 ? »[8] de Tim O'Reilly[9], et où il a été établi une liste de principes pour tenter de clarifier quelque peu ce concept et d'en faire une idée concrète. Ainsi, les sept principes clés exprimés lors de cet article sont :

- ***Penser le Web 2.0 comme une plateforme de services.***

  Cette première règle du Web 2.0, fait que désormais le web doit être vu non pas comme un réseau d'ordinateurs mais plutôt en tant que *plateforme*, on passe alors d'une collection de sites, à une plateforme informatique à part entière, fournissant des applications web aux utilisateurs. Ainsi des éléments comme le « *tagging* », principe ou l'on peut identifier des individus sur des photos, images ou vidéos, des blogs Ou encore le téléchargement et l'échange de fichiers Pair à Pair sont les valeurs et concepts qui font partie intégrante de cette plateforme 2.0.

- ***Émergence d'une Intelligence Collective***

  La deuxième caractéristique est de tirer parti de l'intelligence collective, c'est justement le principe adopté par l'encyclopédie en ligne *Wikipedia* et d'autres sites

8. http://oreilly.com/web2/archive/what-is-web-20.html
9. fondateur d'O'Reilly Media

similaires. Pour Tim O'Reilly : « *l'implication des utilisateurs dans le réseau est le facteur-clé pour la suprématie du Web 2.0 ».*

- ***La richesse est dans les données***

  L'importance des bases de données est une autre notion clé du web 2.0, car toutes ces applications exploitent des bases de données qui s'enrichissent en permanence des contributions et expériences des internautes. Ainsi le web 2.0 marque l'avènement des contenus générés par les usagers du Web.

- ***La longue traine***

  Selon Tim O'Reilly le service s'améliore quand le nombre d'utilisateurs augmente. Le Web 2.0 met ainsi à profit l'effet de la « longue traîne » reflétant un principe économique plus connu sous le nom de loi des « 20/80 » et qui dans le cas du Web 2.0 signifie que ces utilisateurs disposent de données uniques, difficiles à recréer, et dont la richesse s'accroît avec l'augmentation du nombre des utilisateurs**.**

Le cinquième point repose sur des principes plutôt techniques au niveau des modèles et technologies de programmation, en considérant les internautes comme Co-développeurs des applications, quant au sixième il traite d'une préalable évolution du web sur les appareils mobile, ceci semble avoir aujourd'hui considérablement évolué notamment avec le développement des *Smartphones*, telles que *Iphone, BlackBerry*, qui nous permettent d'avoir accès, n' importe où, à des applications en fonction de nos besoins immédiats.

Enfin la dernière caractéristique développée, fut celle de l'enrichissement des interfaces utilisateurs, pour les rendre plus aisées afin d'avoir un aspect ergonomique proche des interfaces PC habituelles et de permettre un usage plus souple du web, pour tout un chacun.

2.2. Outils, Services et Pratiques caractéristiques du Web 2.0

La Figure 2.1, présente sous la forme d'une représentation graphique appelée « nuage de tags », les mots-clefs associés au terme « web 2.0 ». Cette carte a l'avantage de représenter simplement par la variation de la taille, des mots et parfois de leur couleur, la distance sémantique entre des termes connexes. Ces mots-clefs peuvent être répartis en 2 catégories qui concernent :

Figure 2.1. Nuage de Tags du Web 2.0 (d'après Markus Angermeier)

1. Les évolutions techniques notamment en matière de langages de programmations, protocole etc, permettent ainsi d'offrir aux utilisateurs des possibilités de développement de contenus riches en fonctionnalités, tels qu'Ajax, Ruby On Rails, REST, XML, etc.

2. L'évolution des usages notamment avec l'apparition de nouveaux standard et outils favorisant l'aspect communicationnel et collaboratif des usagés du web ; ce qui donne ainsi lieu à de profonds changements dans la façon dont est généré le contenu sur le Web.

Nous proposons dans ce qui suit d'aborder le Web 2.0 sous ses deux aspects, nous recenserons dans un premier volet ces outils permettant la création ou le déploiement du contenu Web 2.0 proprement dit et dans un deuxième volet tous les nouveaux usages et services qui font de celui-ci, un web participatif et collaboratif.

- **VOLET 1 : TECHNOLOGIES DU WEB 2.0**

2.2.1. Rich Internet Application - RIA

La notion ***R**ich **I**nternet **A**pplication* met notamment en valeur le septième principe qu'avait décrit Tim O'Reilly à travers son article « Qu'est ce que le Web 2.0 ». Ainsi un client riche web abrégé RIA, représente toute application Web qui offre des caractéristiques similaires aux logiciels traditionnellement installés sur ordinateur ; car l'interactivité et la vitesse d'exécution sont particulièrement évoluées dans ce genre d'applications Web.

Ainsi une RIA peut être soit :

- Exécutée sur un navigateur Internet sans aucune installation requise.
- Exécutée localement dans un environnement sécurisé appelé sandbox[10].

Dans ce genre d'application, le modèle de page en page n'existe plus. A titre d'exemple, un bouton de formulaire ne va pas forcément recharger toute la page mais simplement avoir une influence sur une partie de la page ou charger une image et de ce fait, offre plus de souplesse aux utilisateurs.

Une autre catégorie d'applications est RDA pour *Rich Desktop Application, qui cette fois a pour rôle d'*apporter ce que l'on retrouve dans le web, sur le bureau de notre ordinateur. Cela permet par exemple l'intégration de Widgests[11] offrant des services web tout en se passant du navigateur internet, ou encore l'application peut tout aussi bien fonctionner en mode déconnecté, cela influe énormément sur l'augmentation du taux d'utilisation Web et permet ainsi de fidéliser l'internaute.

Cependant, l'émergence de ce genre d'applications, est fortement due à l'existence de technologies ayant la possibilité de réunir en même temps les facteurs de puissances, souplesse dans un même outil. Parmi eux on retrouve AJAX.

2.2.2. AJAX - Asynchronous JavaScript And Xml –

Cette méthode de développement Web permet d'économiser de la bande passante, en ne rechargeant pas une page entière alors que seuls certains éléments ont besoin de l'être, mais en ne rafraîchissant que ces éléments de la page. Cela permet de produire des contenus dynamiques et réactifs en ne nécessitant que peu d'appels au serveur hébergeant la page. AJAX est une combinaison de plusieurs technologies : HTML, CSS, Javascript, XML…De nombreux sites web 2.0 sont développés en AJAX.

2.2.3. REST

L'architecture REST pour *Representational State Transfer,* est un concept fondateur de l'évolution Web 2.0. Contrairement aux autres technologies du web, REST n'est ni un standard, ni un format ou encore protocole, car il n'existe pas de spécifications de REST connue à ce jour. Il s'agit de principes d'architecture ou d'une manière de construire des applications distribuées. Cette architecture a pour but d'être plus simple et plus efficace que celles reposant sur SOAP et les Web Services. REST utilise les URI comme syntaxe universelle pour adresser les ressources, HTTP pour transmettre les méthodes et leurs paramètres, des hyperliens pour représenter à la fois le contenu des informations et la transition entre états de l'application, ainsi que les types MIME[12] pour l'identification des ressources.

10. Processus *informatique* verrouillé, où se déroulent des programmes en circuit fermé et clos.
11. Window & Gadget.
12. Standard qui étend le format des courriels pour supporter des textes en différents codage de caractères autres que l'ASCII.

L'avantage de ce style architectural est l'absence de gestion d'état du client sur le serveur, ce qui conduit à une consommation de mémoire inférieure et donc une capacité plus grande de répondre à un grand nombre de requêtes simultanées. A titre d'exemple, l'architecture d'*Amazon*[13] et *Flickr*[14] est bâtis sur REST.

2.2.4. Ruby On Rails

Ruby on Rails est un Framework Web libre écrit en Ruby. Il suit le motif de conception Modèle-Vue-Contrôleur. Il permet de créer des applications dignes du Web 2.0 le plus rapidement possible. Il ajoute aussi un grand niveau d'abstraction dans la programmation de l'application, grâce à un ensemble de fonctions de haut niveau permettant de se concentrer surtout sur les fonctionnalités plutôt que sur la mécanique autour de ces fonctionnalités.

2.2.5. Les systèmes d'édition Wiki

Les wikis sont des sites Web permettant la création et l'édition collaborative de contenus de manière simple. Ils reposent généralement sur un ensemble de pages éditables et organisées en catégories. Ils sont devenus le symbole de l'interactivité promue à travers le Web 2.0. L'un des principes fondateurs des wikis, qui constitue également le principal facteur de leur popularité, est leurs simplicités d'utilisation. Les wikis sont créés et maintenus grâce à des systèmes spécifiques de gestion de contenus, les moteurs de wiki. De nombreux langages, appelés les wikitexts, ont vu le jour afin de permettre la structuration, la mise en pages et les liens entre les articles. Chaque système dispose généralement de son propre wikitext.

***Définition** : Un wiki est un système de gestion de contenu de site web rendant ses pages Web librement modifiables par tous les visiteurs y étant autorisés. Les wikis sont utilisés pour faciliter l'écriture collaborative de documents avec un minimum de contraintes. Ils ont été inventés en 1995 par Ward Cunningham, pour une section d'un site sur la programmation informatique qu'il a appelée WikiWikiWeb. Le mot « wiki » vient du redoublement hawaïen « wiki wiki », qui signifie « rapide ». Au milieu des années 2000, les wikis ont atteint un bon niveau de maturité ; ils sont depuis lors associés au Web 2.0. Créé en 2001, Wikipédia est devenu le site web écrit avec un wiki le plus visité.* (Wikipédia)

Alors que *Wikipédia* est un exemple très connu de Wiki, nous voudrions présenter une initiative moins connue nommée GoogleKnol[16] qui s'intéresse aux articles d'experts et qui reflète bien l'aspect collaboratif et participatif du Web 2.0. GoogleKnol a été lancé officiellement en Juillet 2008. « Knol » signifie une unité de «knowledge» (connaissance) et désigne un article de référence sur un sujet. Bien que Wikipédia et GoogleKnol puissent sembler très proches (tout le monde peut écrire des articles), GoogleKnol se différencie en

13. www.amazon.fr
14. www.flickr.com
15. http://rubyonrails.org/.
16. http://knol.google.com

visant plus particulièrement la participation de professionnelles du domaine, ce qui fait toute la différence avec les autres systèmes de Wiki.

- **VOLET 2 : USAGES ET CONCEPTS DU WEB 2.0**

2.2.6. Les technologies de Syndication

Le Web 2.0, c'est aussi la mutualisation des contenus, l'échange standardisé d'informations entre sites, aux travers de formats de flux. Les deux principaux formats à ce jour sont RSS et Atom, tous deux employés à travers d'un service de brèves et post de blogs. L'abonnement à différents flux émanant de sites préalablement choisis permet ainsi de recevoir sous la forme d'un flux ininterrompu les nouveaux contenus produits sur ces plateformes.

2.2.6.1. La Syndication de contenu « RSS »

Créée par Netscape en 1999, la syndication de contenu permet la transmission selon le schéma XML de tout ou partie du contenu d'une page Web sous la forme d'un flux de donnée. Ainsi toutes modifications apportées à un site Web engendrent un flux de données qui peut être récupéré par un abonné via différents « agrégateurs ». L'agrégation consiste à récupérer puis à afficher des flux d'informations, par l'utilisation locale d'un logiciel client comme : *Google Reader* ou *Netvibes*.

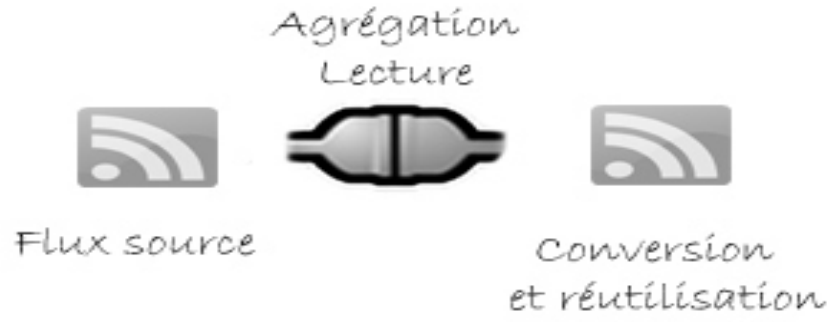

Figure 2.2. Illustration du processus lié à la syndication de contenus

Comme l'illustre la Figure 2.2, les informations récupérées sont analysées, conservées et peuvent être redistribuées par le biais des plateformes de Social Bookmarking, par leur republication sur différents blogs ou instantanément communiquées à l'ensemble de la communauté via les services de micro-blogging. Les flux RSS permettent ainsi de liée des relations au sein des espaces informationnels, contrairement à ce qui ce faisait avant ou l'usager devait faire venir à lui l'information par une démarche active de recherche dite : « info-pull ».

2.2.6.2. Atom

Après quelques problématiques liées à RSS, notamment le fait qu'il était arrivé à sa dernière version définitive; Atom[17] est alors apparus comme une alternative à RSS, basée sur le protocole REST de publication de contenu. Son objectif est de fournir un format équivalent, avec en plus l'utilisation de nouvelles balises XML.

2.2.7. Les Mashups

Le terme « *mashup* » désigne le mélange de deux ou plusieurs éléments pour en créer un nouveau. Sur le Web, un Mashup implique deux ou plusieurs applications web dont le croisement des données va produire un nouveau service qui repose sur l'intelligence collective des internautes et des développeurs. En effet, il peut être à son tour enrichi de contenus générés par les utilisateurs et permet au producteur d'un contenu de le proposer à la communauté des développeurs qui pourront à loisir le formater et le valoriser sous une autre forme. Ainsi le mashup participe intimement au succès du Web 2.0

L'élaboration d'un mashup repose sur une ou plusieurs API (Application Program Interface), ouvertes, librement accessibles et misent à disposition par l'éditeur d'un site web, mais aussi l'objet XMLHttpRequest et AJAX du côté client. Le développeur pourra utiliser cette interface comme une clef d'accès pour récupérer du contenu et interroger des bases de données distantes, à noter aussi que les web services, représentent la base de tout les mashup.
Les API *Google Maps, yahoo* ou d'*amazon* sont les plus utilisées sur le Web, malgré le nombre important de kits de programmation aujourd'hui proposés par les éditeurs de sites.

2.2.8. Blog, Blogosphère et Blogroll

Un *Blogs* est un journal personnel disponible sur le web. Il peut être tenu par un particulier, un chercheur, un journaliste, un salarié d'entreprise ou par un groupe de personne (entreprise, collectivité). Le blog a pour objet de diffuser des billets, généralement liés à l'actualité, et présentés par ordre chronologique. Une *Blogosphère* est la « biosphère des blogs » qui désigne la communauté des blogs, ou encore l'ensemble des auteurs de blogs. Tandis qu'un *Blogroll* est une liste de liens vers d'autres blogs, présentés par l'auteur d'un blog. On peut syndiquer sur une même page des billets venant de ces blogs via les formats RSS.

2.2.9. Le Micro-Blogging

Le *Micro-blogging* est un service simple et rapide de publication de messages. On parle de blogging parce que les personnes les plus actives y sont souvent les bloggeurs. Beaucoup d'usages peuvent en être faits, certains préviennent d'un nouveau post sur leur blog, certains y voient un média d'informations, d'autres publient seulement leurs émotions ou des instants importants de leur vie qu'ils ont envie de partager, d'autres encore l'utilisent à des fins

17. http://tools.ietf.org/html/rfc4287.

politiques, enfin les comportements sont nombreux. Il existe actuellement plusieurs services de micro-blogging et en tête, on retrouve *Identi.ca*[18] et *Twitter*[19]. Twitter est historiquement le premier à être apparu, *identi.ca* a suivi. Ce dernier est une plateforme libre contrairement à Twitter qui est propriétaire.

Le micro-blogging se présente donc par un aspect très léger, pas de profil publiée, pas d'applications partagées, pas de communautés ou de groupes, pas de photos ou autre, seulement une page où l'on consulte en une liste les messages des personnes que l'on décide de suivre. Ces messages sont ordonnés par date de publication et affichées dès qu'ils ont été publiés en temps réel. Le contenu varie en fonction des gens, et on retrouve les suiveurs et les suivis. On peut bien sûr être les deux à la fois, donc le suiveur ou *follower* est celui qui s'abonne aux messages du *suivi* ou *followed*. Sur Twitter on retrouve les termes *abonnement / abonné* respectivement pour *following / follower*.

Le principe de Twitter est que chaque usager qui a préalablement procédé à l'ouverture d'un compte, poste un message, depuis le site web du service ou depuis l'une des 1300 applications tierces, et chaque message posté est instantanément transmis à l'ensemble des usagers qui « suivent » (abonnés) l'émetteur du message, qui lui même suit d'autres usagers qui engendre ainsi non pas un maillage de documents comme sur le web classique, mais un maillage dont les nœuds sont des individus.

A noter que l'un des intérêts du micro-blogging est tout d'abord la discrétion et l'intégration du service dans les outils que l'ont utilise au quotidien, pour les avoir tout le temps sous la main. Ainsi, les micro-blogs sont utilisables soit directement sur les pages officielles des services identi.ca et twitter, intégré à son navigateur via des extensions (Identifox, Twitterfox) ou carrément sur services de réseaux sociaux implémentant ces clients de micro-blogging tels que Facebook, Netvibes, Twhirl, Tumblr et FriendFeed.

L'ancienneté et le nombre d'abonnés confère à *Twitter* la première place des sites de Micro blogging, il existe cependant des plateformes concurrentes qui se distinguent par l'ajout de fonctionnalités complémentaires comme l'illustre le (tableau 2.1) ci-dessous :

Plateforme	***Contenu***	***Spécificité***
Identi.ca	Texte	Open Source
TumblR	Texte/Photo	Photo blogging
Jaiku/Google	Texte/Icones	Géo Localisation
Hictu	Texte/Vidéo	Vidéo blogging

Tableau 2.1. Liste des principales applications de micro-blogging[20].

18. www.*Identica.com*
19. www.twitter.com
20. http://asterisq.com

2.2.10. Les Tags et les Folksonomies

Les Tags sont des étiquettes que les usagers peuvent joindre virtuellement sur un document digital : article ou photo par exemple. Ils peuvent ensuite s'en servir pour classer le document ou le partager avec d'autres. Les tags sont aussi très populaires sur *Flickr*, un site sur lequel on peut garder ses propres photos de façon privée ou publique.

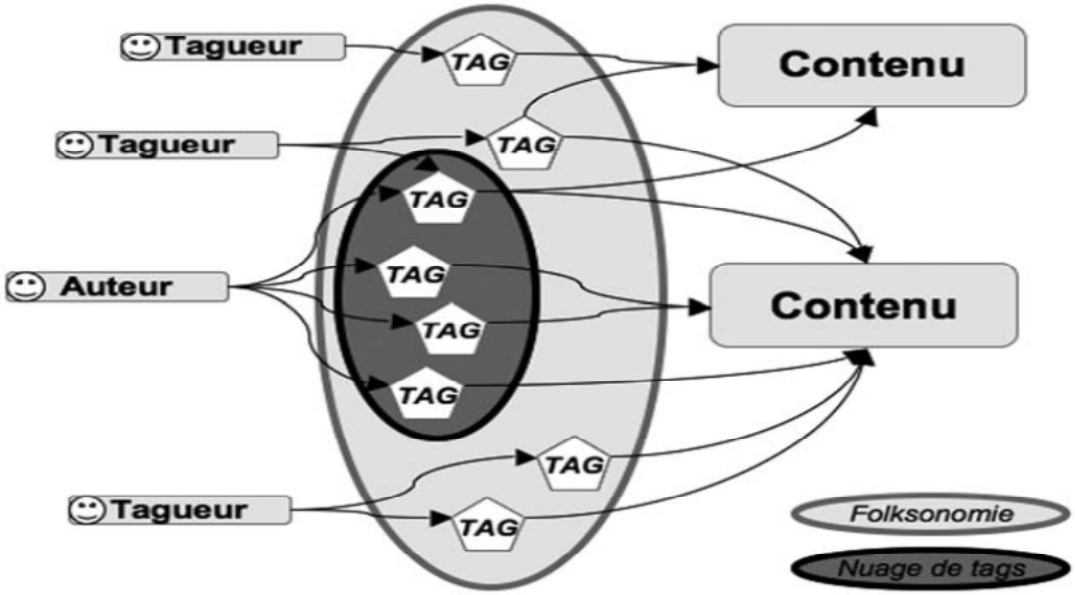

Figure 2.3. Graphique de composition de Tags

On retrouve encore les Nuages de Tags, appelé aussi « Tag Cloud » qui sont une autre façon de présenter les tags correspondant à un site et qui consiste à regrouper sur un même espace les mots en question en leur attribuant une taille variable suivant la fréquence d'utilisation. C'est très commode pour trouver instantanément ce qui est le plus populaire à un moment donné.

Figure 2.4. Nuage de Tags (Extrait de Wikipedia)

La Folksonomie quant à elle est à l'opposé des formes traditionnelles de classification des connaissances conçues à priori, cette organisation est faite par les gens, ne s'interrompt jamais et progresse avec le temps. Les agrégats d'individus qui constituent des communautés de fait, reposent sur le principe de la « *folksonomy* », selon Wikipedia c'est un *« système de classification collaborative décentralisée et spontanée »*.

L'intérêt des Folksonomies est lié à l'effet communautaire, car pour une ressource donnée sa classification est l'union des classifications de cette ressource par les différents contributeurs. Ainsi, partant d'une ressource, et suivant de proche en proche les terminologies des autres contributeurs il est possible d'explorer et de découvrir des ressources connexes. C'est pour cette raison que le concept de Folksonomie est considéré comme faisant partie intégrante du Web 2.0.

En résumé, les caractéristiques de la Folksonomies sont :

- Rendre l'information accessible à tous.
- Utilisation facile, multidirectionnelle et bidirectionnelle.
- Contribue à la formation du Web.
- Facilité de production car tout le monde peut le produire facilement. Aucune compétence n'est requise.
- Permet à l'usager d'indexer des documents ou des informations et de les retrouver grâce à une classification des données à l'aide de mots-clés.
- Permet d'accroître et de rendre plus rapide la diffusion de l'information.
- Les utilisateurs sont libres d'y choisir leurs propres mots-clés. La folksonomie est réellement centrée sur l'utilisateur. Il peut en faire une utilisation personnelle, professionnelle etc. De cette façon, l'utilisateur peut devenir plus structuré et cela lui permet d'organiser toutes ses informations.
- Permet d'observer les habitudes des internautes lorsque ceux-ci rentrent leurs propres termes quant-aux noms d'étiquettes.

2.2.11. Le Social Bookmarking

Le Social Bookmarking ou signet social est une façon pour les internautes de stocker, de classer, de chercher et de partager leurs liens favoris. Dans un système ou réseau de *bookmarking* social, les utilisateurs enregistrent des listes de ressources web qu'ils trouvent utiles. Ces listes sont accessibles aux utilisateurs d'un réseau ou site web. D'autres utilisateurs ayant les mêmes centres d'intérêt peuvent ainsi consulter les liens par sujet, catégorie, étiquette ou même de façon aléatoire. En dehors des favoris web, on peut trouver d'autres services spécialisés sur un sujet particulier (alimentation et boissons, livres, vidéos, commerce en ligne, cartographie...). Ce phénomène illustre bien la dynamique du Web 2.0 par :

1. Son orientation participative et collaborative, par la mutualisation de ressources individuelles.
2. La mise à disposition d'API, permettant une exploitation des données et leur reformulation dans de nouvelles applications (Mashup).

Ainsi, ce phénomène du web 2.0, permet de refléter en temps réel les opinions de millions d'individus et leurs appréciations individuelles contrairement aux traitements automatisés pratiqué par les moteurs de recherche.

2.2.12. P2P

Le Pair-à-Pair (P2P) désigne les outils et les réseaux permettant des échanges de fichiers directement entre utilisateurs, et ne nécessitant pas de machine centrale. Cette pratique s'est notamment développée avec l'échange de fichiers musicaux au format MP3 ou de films au format DivX.

Napster fut parmi les premiers réseaux P2P à succès. Suite à sa fermeture par la justice, d'autres protocoles et réseaux ont vus le jour, tel que BitTorrent etc. Dans ce réseau les utilisateurs ne sont pas reliés un à un mais plusieurs à plusieurs. Ceci signifie que l'échange ne
s'effectue pas simplement d'un utilisateur à un autre, mais qu'une centaine d'utilisateurs peuvent s'échanger un même fichier en même temps, chacun envoyant aux autres les bouts qu'il possède. Cette technique accélère fortement le téléchargement de gros fichiers.

2.2.13. Long Tail (La longue traîne)

Le phénomène de Longue Traine s'inspire d'une théorie ayant fait ses preuves en statistique, appelé Loi des « 80/20 » et qui s'applique maintenant en marketing Web 2.0, c'est en quelque sorte la possibilité qu'offre le Web 2.0, de sortir de la production et de la consommation de masse pour servir les marchés à faible participation.

Le phénomène de longue traîne fut utilisé pour la première fois par des sites de e-commerce, elle s'est ensuite utilisé dans le domaine du référencement et c'est maintenant devenu son domaine d'usage le plus courant. Dans ce cadre, il désigne le fait qu'un partie importante du trafic en provenance des moteurs, puisse provenir d'un grand nombre de mots clés distincts, apportant chacun peu de visites, le service *Google AdSense* tire parti justement de ce phénomène. Aussi, dans le domaine des liens commerciaux, le fait d'enchérir sur les mots de la longue traîne peut permettre de bénéficier de niveaux d'enchères moins élevés car la concurrence y est moins forte et ainsi en profiter pour faire de bonnes affaires comme c'est le cas sur *eBay*, ainsi que son analogue algérien *Ouedkniss* qui cependant n'est pas de même envergure d'eBay bien entendu.

2.2.14. Les Réseaux Sociaux

Réseaux sociaux ou encore Social networks sont des noms que l'on ne cesse d'entendre depuis quelques temps, en effet à l'origine ces techniques était plus convoités par une sphère de jeunes bloggeurs et étudiants, permettant ainsi de renforcer et divertir leurs communautés virtuelles qui étais généralement proche géographiquement, alors que maintenant, nous entendons par réseaux sociaux les sites où les liens entre membres et la création de groupes d'utilisateurs, parfois appelés « communautés », sont largement exploités. Il existe différents types de réseaux sociaux, ceux réservés à un usage généraliste tel que Facebook, où à usage spécialisé comme MySpace qui est a été conçu à l'origine pour les communautés musicales, et d'autres encore à visée professionnelle tel que LinkedIn ou Viadeo.

Alors qu'un outil comme Wikipédia permet la structuration d'une communauté autour d'objets communs (les articles encyclopédiques), un réseau social permet l'apparition digitale de réseaux existants dans le monde physique. On pourrait considérer que le contenu crée par les utilisateurs sur un réseau social comme Facebook est l'ensemble des traces digitales de sa vie virtuelle.

Figure 2.5. Illustration des différents réseaux sociaux.

D'un point de vue purement informatique, Les réseaux sociaux représentent des structures sociales dynamiques se modélisant par des sommets et des arêtes. Les sommets désignent généralement des gens et/ou des organisations et sont reliées entre eux par des interactions sociales. L'analyse des réseaux sociaux, basée sur la théorie des réseaux, l'usage des graphes et l'analyse sociologique représente le domaine étudiant ce type de réseaux.

Cependant, l'émergence des réseaux sociaux et ces multiples interactions sont à l'origine de la désorganisation des données. Pour remédier à cela le concept de Métadonnées semble être pour le moment la meilleure solution, puisque elle permet de garder des interactions tout en structurant les données, ce qui représente le principe clé du web sémantique. Mais les Web services jouent eux aussi un rôle important en permettant l'exploitation de ces données à l'image d'un outil. La fusion de ces deux concepts ou ce que

certains nomment le Web 3.0, conduira à une meilleure exploitation des données issues des réseaux sociaux.

2.3. Limites du Web 2.0

2.3.1. La Surcharge d'information

Au cours de ces six dernières années, le nombre de service Web 2.0 a considérablement augmenté, la facilité d'utilisation de ces services ont permis la prise de contrôle de l'information par les utilisateurs. En effet, n'importe quel internaute peut aujourd'hui collaborer, partager des informations, et tout ceci sans connaissances spécifique. Cet effet d'appropriation du Web par les internautes a évidemment des conséquences, notamment en termes de quantités de données disponibles en ligne. Cette quantité de données hétérogènes (Pages web, Flux RSS, notifications des réseaux sociaux, billets de blog, commentaires, Newsletters …) présentes sur la toile s'en est trouvée multipliée et continue de l'être jour après jour, d'ailleurs selon une évaluation de l'Internet World Stats[21] réalisée en 2010, il y aurait plus de 1,9 milliards d'internautes dans le monde pouvant consulter environ 109,5 millions de sites Web opérationnels et 25.21 milliards de pages.

Or, devant cette surabondance d'informations, l'utilisateur devient incapable de gérer cette quantité de flux d'informations qui lui parvient quotidiennement via les différents agrégateurs de flux. Il devient ainsi pénible de sélectionner les informations qui correspondent au mieux à ses attentes. Et pire encore, l'utilisateur se retrouve face à une information de mauvaise qualité (beaucoup de redondances, information inutile, ressources non intéressantes) ce qui implique une énorme perte de temps lors des recherche d'information pertinente, ceci représente ainsi l'une des préoccupations majeurs de la communauté Web 2.0.

Face à cette problématique de surabondance de données et d'informations, on peut finalement se poser les questions suivantes : *« **(i)** Quelles sont les solutions les plus utilisés à l'heure actuel pour palier à ce problème de parcours de données et de contenus ?, **(ii)** répondent elles parfaitement aux exigences des utilisateurs en terme de pertinence des résultats et quel serais le moyen de remédier à ce manquement ? **(iii)** profitent elles réellement de l'aspect collaboratif du web 2.0 ? ».*

Pour répondre à ces questions, nous nous sommes principalement intéressés aux domaines de recherche touchants à cette problématique à savoir, la *Recherche d'Information* et le *Filtrage d'Information*.

21. www.internetworldstats.com

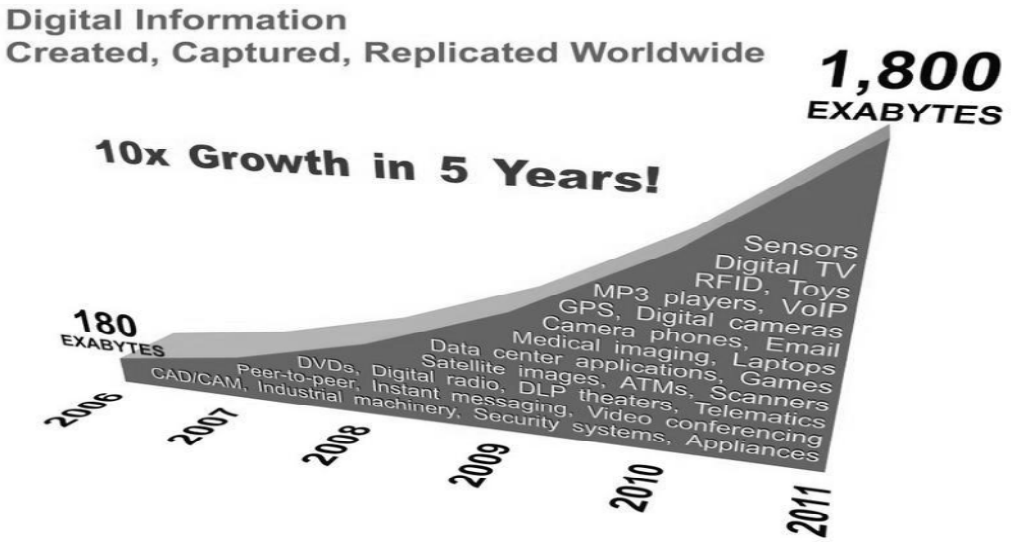

Figure 2.6. The Diverse and Exploding Digital Universe (d'après IWS)

- **Discussions autour des solutions existantes**

➢ **Les Moteurs de Recherche**

L'une des voies remédiant au problème de surcharge informationnelle, et ayant apparue bien avant l'arrivée du Web 2.0 est la recherche d'information et ces fameux moteurs

de recherches. Ces derniers ont pour rôle d'explorer et de parcourir le Web, afin d'indexer les pages qui y sont publiées. Cette indexation consiste en l'extraction de mots-clés, considérés comme significatifs, et représentant le contenu des pages. L'objectif de ces moteurs de recherches est de proposer des informations correspondantes aux équations de recherche formulées par les utilisateurs (sous forme de mots-clés).

La dernière décennie a été marquée par une évolution considérable des moteurs de recherche dont *Google* est devenu le plus populaire. Celui-ci est basé sur une indexation automatique des contenus notamment grâce à son algorithme nommé Page Rank[22], calculant la popularité de chaque page Web en fonction des liens hypertextuels pointant dessus. De manière simplifiée, si beaucoup de sites sur le Web renvoient vers un site A alors il y a des chances que ce site A soit intéressant. Ainsi lorsque les machines de Google parcourent le Web pour indexer les contenus, elles comptent en même temps le nombre de liens partant et arrivant sur chaque page. Ceci permet de déterminer une popularité puis un ordre entre les pages.

Cette simplicité d'utilisation a permis à Google et ses concurrents d'être utilisés par une grande majorité d'internautes. Cependant, ces moteurs de recherche basés sur l'indexation automatique présentent certaines difficultés, à titre d'exemple pour déterminer des caractéristiques implicites comme la qualité informationnelle d'un document, ces moteurs

22. Algorithme d'analyse des liens, inventé par Larry Page.

comptent les liens entre pages mais ne jugent pas directement le contenu. S'il n'y a pas assez de liens entre des documents, ces moteurs seront incapables d'ordonner statistiquement les documents. Aussi, ces moteurs étant encore très liés aux termes des documents, ils offrent peu d'aide à un utilisateur ne connaissant pas un domaine et ses mots clefs et ne fournissent pas de listes de concepts principaux d'un domaine à l'usage des non-experts, pour faciliter leur exploration.

De ce fait, la qualité et la pertinence des informations proposés par les moteurs de recherche sont notamment conditionnées par la précision des équations de recherche des utilisateurs. Ainsi pour répondre à la question ***(ii)***, on s'est intéressé à une des évolutions technologique du Web : Le Web Sémantique.

En effet, le Web sémantique se présente comme une solution complémentaire à la recherche de contenus, et l'existence des ontologies permet d'indexer non pas selon des mots-clefs mais selon des concepts reliés entre eux. Les moteurs de recherche s'appuyant sur les ontologies peuvent donc déduire que si un utilisateur recherche à titre d'exemple une « Mémoire vive », il peut être utile de lui fournir aussi les documents indexés par « RAM ». Grâce aux ontologies, les machines disposent ainsi d'une connaissance du domaine pouvant améliorer les résultats des recherches.

Aussi, pour réponde à la question ***(iii)***, on remarque qu'à l'heure actuel il existe très peu de moteurs de recherche qui exploitent les techniques du web collaboratif pour profiter de la ré-indexation et de l'enrichissement des contenus par les usagers eux mêmes. Pourtant ces techniques pourraient bien apporter des solutions dans le cadre de la recherche de données, En effet il serait intéressant d'opter pour des solutions qui permettraient (en plus de celles existantes) une indexation prenant en compte le point de vue de l'usager. Ce dernier en consultant le document peut construire des informations à valeur ajoutée sur le contenu consulté, il s'agira par là de nouvelles informations qui enrichissent le contenu et qui permettent une certaine visibilité sémantique et une facilité de lecture par des annotations, des liens vers d'autres contenus, de nouveaux descripteurs, et même des résumés propres à l'usager. Le principe de *Folksonomie* cité auparavant en est un bon exemple malgré la présence de légères imprécisions diminuant la qualité des résultats et que nous discuterons plus tard.

Une autre solution serait d'essayer de fusionner ces deux approches (Web sémantique et collaboratif) pour en quelque sorte essayer de mutualiser leurs avantages et compenser leurs limites. D'ailleurs différentes solutions explorent la complémentarité entre ces deux approches dans le cadre de la recherche d'information sur le web. L'expression « Web 3.0 » dans [Mika, 2007] a été souvent employée pour décrire cette hybridation. Celle-ci peut alors s'effectuer à titre d'exemple en proposant des solutions [Section 2.3.2] qui récupèrent les tags hétérogènes des utilisateurs puis tentent de les transformer en des concepts structurés, que les machines vont pouvoir plus facilement échanger et traiter. Ces solutions s'appuient généralement sur les technologies du Web Sémantique, une autre alternative serais aussi de

proposer des outils sémantique pour aider les utilisateurs à percevoir et négocier leurs différents points de vue, et par conséquent leurs recherches.

En conclusion, on voit bien que le Web Sémantique et le Web collaboratif offrent des possibilités prometteuses pour améliorer la recherche de donnée et de contenus pertinents. Cependant, le rapprochement de ces techniques n'en est qu'à ses débuts, nous aborderons certains travaux à ce sujet dans de prochains points.

- ***La Personnalisation***

Une autre voie, qui elle s'est largement développée avec l'arrivée du Web 2.0, est la personnalisation. Par Personnalisation, on entend par là l'adaptation des pages web pour un utilisateur en particulier. Le but est de moduler le web afin d'aider les internautes à accéder, le plus simplement et rapidement possible, aux ressources qu'ils désirent. Cette adaptation peut être faite manuellement par l'utilisateur, comme c'est le cas avec la majorité des portails communautaires ou chaque utilisateur peut organiser sa page, en déplaçant, ajoutant, supprimant ces contenus et ainsi se simplifier l'accès vers les données qui l'intéressent le plus. Ces enjeux liés à la satisfaction des attentes des utilisateurs et à leur fidélisation constituent les objectifs principaux de la personnalisation de l'accès à l'information. La personnalisation est un axe de recherche qui suscite l'intérêt et l'engouement de nombreux chercheurs. Plusieurs approches ont été ainsi proposées, La personnalisation automatisée se base principalement sur la recommandation et plus précisément sur le filtrage d'information.

Ces systèmes de recommandation sont donc une des solutions à cette personnalisation automatisée. Ils sont devenus, à l'instar des moteurs de recherche, un outil incontournable pour aider les gens à surmonter la surcharge d'information. Un système de recommandation est un type spécifique du filtrage d'information qui présente des éléments qui sont susceptibles d'intéresser un utilisateur. Les travaux dans le domaine de recommandation se focalisent à l'heure actuel sur la conception et le développement des algorithmes. Il existe plusieurs types d'algorithmes de recommandation, les deux approches les plus utilisées sont le filtrage collaboratif et le filtrage de contenus. Une définition plus formelle de la recommandation est donnée par Adomavicius [Adomavicius et al., 2005].

Définition. Soit U l'ensemble de tous les utilisateurs, soit I l'ensemble de tous les éléments qui peuvent être recommandés, soit R un ensemble ordonné et soit $f : U \times I \rightarrow R$ une fonction qui prédit l'intérêt que portera l'utilisateur $u \in U$ à l'item $i \in I$. Alors pour chaque utilisateur $u \in U$, le système de recommandation sélectionne l'élément $i' \in I$ qui maximise l'intérêt de u :

$$\forall u \in U, i'_u = argmax_{i \in I} f(u, i)$$

L'intérêt d'un utilisateur pour un élément (la fonction f (u, i)) est généralement représenté par une note indiquant l'appréciation que l'utilisateur porterait sur l'élément. Afin de deviner cet intérêt, des connaissances sur l'utilisateur en question sont néanmoins nécessaires. Les goûts connus des utilisateurs sont généralement caractérisés par leurs appréciations portées sur les contenus déjà consultés. Ces informations sont regroupées dans une matrice appelée « matrice d'usages ». Le tableau ci-dessous présente un exemple fictif de matrice binaire contenant des informations de type « l'utilisateur u a apprécié /n'a pas apprécié l'élément i ». Elles peuvent également se mesurer sur un nombre plus élevé de classes « à mis 1/2/3/4/5 étoiles », etc. Une fois la matrice d'usages construite, l'objectif du moteur de recommandation est de deviner les connexions utilisateur-élément manquantes. En d'autres termes, on demande à l'outil de remplir les cases vides C_{ui} de la matrice en évaluant si l'élément i intéressera l'utilisateur u ou non. Comme nous l'avons précisé ci-dessus, il existe deux approches de filtrage les plus utilisées : le filtrage collaboratif et le filtrage de contenus, cela dépend de la manière avec laquelle l'utilité ou la pertinence éventuelle est calculée ou estimée.

	Elément 1	Elément 2	Elément 3	Elément 4	Elément 5	…
Utilisateur 1	Satisfait			Satisfait		…
Utilisateur 2		N/ satisfait	Satisfait			…
Utilisateur 3		Satisfait				…
Utilisateur 4			N/ satisfait	Satisfait	N/ satisfait	…
…	…	…	…	…	…	…

Tableau 2.2. Exemple de matrice d'usages

On retrouve donc, *les systèmes de filtrage à base de contenu* qui recommandent des documents similaires à ceux que l'utilisateur a déjà appréciés. Ceci est calculé en rapprochant les centres d'intérêt des utilisateurs (introduits de manière explicite à travers un questionnaire par exemple ou de manière implicite à travers la surveillance de son comportement) avec la métadonnée ou les caractéristiques des documents, sans prendre en compte les avis des autres utilisateurs [Rao et al., 2008]. Plus formellement, les documents sont représentés sur un vecteur X = (x1, x2, ..., xn) de n composantes. Chaque composante représente un attribut et peut contenir des valeurs binaires ou numériques. Dans le cas de la recommandation d'article par exemple, les attributs peuvent être la spécialité, l'auteur, l'année de publication, etc. Ce type de vecteurs fait alors office de profil. Une fois les profils construits, l'objectif du moteur est d'évaluer leurs similarités.

Le filtrage basé sur le contenu présente néanmoins un certain nombre de limitations. D'abord il est difficile de faire des indexations pour les objets à rechercher, aussi le système s'appuie sur un profil qui décrit le besoin de l'utilisateur du point de vue thématique et ce profil peut prendre diverses formes. Une autre difficulté est l'effet dit « entonnoir » qui restreint le champ de vision des utilisateurs. En effet, ce type de filtrage est incapable de recommander des documents qui sont différents de ceux que l'utilisateur a déjà vus et évalué, ce qui ne laisse pas de place à des documents pourtant proches mais dont la description thématique diffère fortement.

Quant au *filtrage collaboratif*, il est basé sur l'hypothèse que les gens à la recherche d'information peuvent se servir de ce que d'autres ont déjà trouvé et évalué. Ce type de moteur utilise uniquement les informations contenues dans la matrice d'usages comme données d'entrée. La matrice peut être construite en surveillant les comportements des utilisateurs ou encore en proposant aux utilisateurs de déclarer eux-mêmes leurs avis sur les items qu'ils connaissent. Afin de remplir les cases vides de la matrice. Deux grands axes se distinguent dans la littérature. Les approchés basées sur les plus proches Voisins, appelées aussi approches basées sur la mémoire, et les approches basées sur les modèles. Des hybridations de ces approches existent également.

Concernant les approches basées sur les modèles, elles mettent en œuvre des méthodes issues de l'apprentissage automatique (Machine Learning) comme des méthodes de clustering. Ces méthodes sont généralement performantes mais ont un coût de construction et de fonctionnement plus important que les méthodes basées sur les plus proches voisins [Candillier et al., 2007], [Khoshgoftaar, 2009]. De plus, ces méthodes semblent plus efficaces que les approches basées sur les plus proches voisins uniquement dans le cas de données d'usages clairsemées. Le filtrage par plus proche voisin lui aussi présente des limitations. Un problème du système collaboratif est que sa performance dépend beaucoup de la distribution des évaluations données par utilisateurs. Si dans le système il y a plusieurs articles qui ont été utilisés et évalués par très peu d'utilisateurs, ces articles seraient recommandés très rarement, même si ces utilisateurs ont donné des notes très hautes pour ces articles. Ce problème est connu comme *le problème de parcimonie*. De la même façon, si dans le système il existe des utilisateurs qui ont des goûts très différents en comparaison avec les autres, le système ne peut pas trouver des similarités entre utilisateurs et donc ne peut pas donner de bonnes recommandations.

Ainsi pour éviter cette limitation certains systèmes utilisent la méthode hybride, une combinaison de ces deux méthodes. Plusieurs travaux se sont orientés vers la construction de ce type de méthode [Claypool *et al.*, 1999] ont combiné les recommandations produites par les deux méthodes à base de contenu et collaboratives appliquées séparément. [Pazzani, 1999] a appliqué les algorithmes de filtrage collaboratifs sur une matrice décrivant les préférences des utilisateurs sous forme de mots-clés pondérés au lieu de la traditionnelle matrice des votes. Et enfin [Mellville *et al*., 2002] ont utilisé les recommandations à base de contenu pour compléter la matrice des votes et ensuite appliquer l'algorithme de recommandation collaboratif sur cette matrice.

La croissance exponentielle du Web social et notamment des réseaux sociaux, présente elle aussi des possibilités pour la recherche en système de recommandation. Ainsi en réponse à la question ***(iii)***, on remarque que l'aspect interactif et participatif (notamment des réseau sociaux) qui caractérise le web 2.0 semble bénéfique, car il crée des nouvelles sources d'informations pour la recommandation, soit de façon explicites (quant les utilisateurs déclarent leurs profils sous forme des centres d'intérêts), ou de façon implicites (ils expriment implicitement leurs préférences via leurs activités et leurs interactions avec les autres dans le réseau) ce qui apporte de la valeur ajoutée aux donnée et permet ainsi d'améliorer les

techniques de recommandation existantes voir même développer des nouvelles stratégies, qui se focalisent sur la recommandation sociale. Ainsi, la question qui se pose est : *« comment peut-on exploiter les modèles sociaux dans les systèmes de recommandation et de filtrage d'informations ? »*

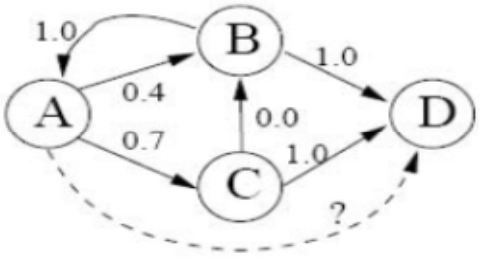

Figure 2.7. Exemple de réseau social avec les scores de confiance [Massa et al.,2004]

Une première idée serait de remplacer la formation des communautés classique sur la base des votes avec celle induite par le réseau social (amis et amis d'amis). Des comparaisons ont été réalisées pour les recommandations collaboratives classiques avec celles faites par les amis sur le système de recommandations (Amazon.com). Les résultats ont montré que les utilisateurs ont préféré celles réalisées par leurs amis. Ceci peut être expliqué par le fait que les amis sont plus qualifiés à les conseiller puisqu'ils sont supposés connaître davantage les préférences des utilisateurs. [Groh *et al.*, 2007] ont pour leurs part mené une étude comparative de la performance d'un algorithme de filtrage collaboratif classique par rapport à un algorithme de filtrage social où la communauté sociale est constituée des amis de l'utilisateur [1] et de leurs amis [2].

$F^{(u)} = \{u_j / \ Q_{ij}=1\}$ [1]

$N_{social}^{(u)} = F^{(u)} \cup \{u_k / u_j \in F^{(u)} : A_{jk=}\ 1\}$ [2]

Avec :

Aij=1 Il existe un arc liant l'utilisateur *i* à *j* dans le réseau social (*i* et *j* sont amis)

La prédiction des votes est calculée en considérant l'union des deux communautés, collaborative et sociale. Après expérimentation, les auteurs ont constaté que l'approche sociale est nettement plus performante dans le cas ou on ne dispose pas de suffisamment de votes. [Ziegler *et al.*, 2007] ont mené une étude expérimentale sur la relation entre la similarité des profils utilisateurs et le concept de *confiance sociale.* Ils ont développé un site web de réseautage social en ligne, FilmTrust, où les utilisateurs peuvent passer en revue des films, les évaluer et gérer leurs profils en utilisant le vocabulaire FOAF(*Friend of a Friend vocabulary*). Les auteurs proposent ensuite un algorithme pour la propagation du degré de confiance sociale à travers le réseau social, TidalTrust [3].

$$t_{i,s} = \frac{\sum_{j \in adj(j) / t_{i,j} \geq \max} t_{i,j} t_{j,s}}{\sum_{j \in adj(j) / t_{i,j} \geq \max} t_{i,j}}$$ [3]

Avec :

$t_{i,s}$ Le degré de confiance entre l'utilisateur actif *i* et la cible *s*
adj(*j*) L'ensemble des utilisateurs adjacents à l'utilisateur *j*
$t_{i,j}$ Le degré de confiance liant *i* à *j*
max Un seuil du degré de confiance

Les résultats ont montré que plus le degré de confiance est grand plus la différence des votes diminue indiquant ainsi une *forte* corrélation entre la similarité et la confiance sociale. [Massa *et al.* 2004], présentent le modèle « Web of Trust » où les utilisateurs définissent un ensemble d'amis à qui ils font *confiance*. Ce modèle à comme entrée la matrice des votes <utilisateurs, documents> et la matrice des scores de confiances entre utilisateurs <utilisateurs, utilisateurs> et produit en sortie une matrice des votes estimés.

2.3.2. Ambiguïté des tags et faible organisation des Folksonomies

Une autre limite du Web 2.0, toujours dans le contexte de la recherche d'information, est due à l'utilisation de tags dans le but d'annoter les différents contenus produits. Si les avantages des tags sont multiples en termes d'annotation, et cela en permettant à l'utilisateur d'adapter les termes à ses souhaits particuliers, cette ouverture contribue à complexifier la recherche d'information. [Mathes, 2004] estime ainsi « qu'une folksonomie représente simultanément une partie du pire et du meilleur dans l'organisation de l'information ». En effet, une folksonomies n'est qu'un amas de tags chaotiques et non organisés. Il devient ainsi difficile d'accéder à l'information si l'on ne se réfère pas directement au tag souhaité et il est encore plus complexe d'étendre ou de spécifier sa recherche. Ainsi, certains pensent que si le gain de temps est considérable en termes de publication, il est perdu en termes de recherche d'information et que la pratique de tagging perd ainsi de son intérêt. Parmi les principales problématiques liées aux pratiques de tagging on retrouve :

- ***Le Problèmes d'ambiguïté***

Un tag peut en effet être associé à plusieurs significations. Par exemple le mot-clé ADE peut correspondre à l'*Algérienne Des Eaux* où à un *Atelier Dépannage Electroménager*, ceci selon le contexte de l'annotation et le contenu annoté. Une recherche sur ce terme récupérera cependant les contenus annotés par le mot-clé quelque soit son sens, induisant un problème d'ambiguïté. Les mots-clés ne portent en effet pas suffisamment de sémantique pour définir par eux-mêmes et sans ambiguïté l'entité qu'ils représentent.

- ***Le Problèmes d'hétérogénéité***

Si un tag peut avoir plusieurs significations, il est également possible que plusieurs tags soient utilisés pour représenter la même entité. C'est là toute l'ambiguïté des systèmes de tags et du choix de ces termes par les utilisateurs eux-mêmes. Cette hétérogénéité est souvent due aux (synonymes, pluriels, variations de case, Multilinguisme ...). Par exemple les tags

ADE, SONAD, EPTO, l'Algérienne Des Eaux et Algérienne_des eaux identifient la même entreprise. Cependant si des systèmes de suggestion ou d'auto complétion peuvent permettre de restreindre cette hétérogénéité, il arrive toutefois qu'elle soit motivée pour des raisons liées à des choix plus personnels (on trouve par exemple sur *Delicious* un certain nombre de tags débutant par _ permettant leur placement en début de liste alphabétique).

- ***Absence d'organisation***

Une dernière limite associée à ces pratiques de tagging est l'absence d'organisation entre tags. Une folksonomie n'est en effet qu'un ensemble de mots-clés désorganisés au sens où aucune relation n'est explicitement définie entre les termes utilisés. Ainsi, bien qu'il puisse exister une relation entre les concepts représentés par différents tags, celle-ci n'est prise en compte à aucun moment. Ces systèmes ne sont ainsi pas capables d'identifier la relation qui existe entre les tags *Web sémantique* et *Web des données* (ou plutôt entre les concepts correspondants), et en conséquence de prendre en compte cette relation au niveau de la recherche d'information et de la navigation. À nouveau, cette absence d'organisation est liée au manque de sémantique qui existe dans des organisations comme les folksonomies.

Une des solutions qui pourrait palier à ses limitations des folksonomies serait de les rapprocher des représentations sémantiquement structurées. Ce processus de structuration permet de transformer ces métadonnées hétérogènes et non structurées que sont les tags, en des métadonnées plus cohérentes et structurées (les ontologies) et de ce fait, facilitant la recherche d'information à base de Tags. Ainsi, nous avons recensés tout au long de la préparation de ce mémoire, de nombreux travaux de recherche qui s'intéressent à ce rapprochement, des travaux qui suivent dans l'ensemble la même problématique de structuration, mais qui présentent tout de même certaines divergences dans leur façon d'aborder ce processus de sémantisation, [Passant, 2009] les a classés en deux grandes familles :

- Des travaux cherchant à identifier ce que l'on appelle « une sémantique émergente » depuis les folksonomies, voire à extraire des modèles taxonomiques ou ontologiques à partir de celles-ci.

- Et les travaux visant à proposer des modèles de représentation pour les tags, les folksonomies et les objets associés (actions de tagging, nuages de tags ...) avec les technologies du Web Sémantique.

Cette première famille analyse a posteriori la structure des folksonomies pour obtenir une représentation de la structure sémantique induite par les communautés, comme première étape à l'élaboration d'ontologies. On retrouve par exemple, [Mika, 2007] qui propose de construire des « Ontologies légères » à partir de l'analyse de folksonomies. Pour cela il propose un modèle « tripartite » des folksonomies où les ressources Web (instances) sont associées par un internaute (acteur) à une liste de tags (concepts) de la manière suivante :

Tagging (*Utilisateur,Ressource, Tag*). Des méthodes d'analyse de réseaux sociaux sont ensuite utilisés pour tisser des réseaux reflétant les liens entre les concepts et en déduire des regroupements de termes, ou des relations de subsumption (relation « is - a »).

[Au et al. 2007] illustre cette approche tripartite (tags, utilisateurs, documents) en désambiguïsant le tag « sf » dans Delicious. En effet, ce tag peut signifier aussi bien « Science Fiction » que « San Francisco » pour les utilisateurs. Pour cela, le système détermine les graphes des documents et des utilisateurs. Il y a un lien entre deux documents si un utilisateur les a taggués avec le même tag, et il existe un lien entre deux utilisateurs s'ils ont tagué un document avec le même tag. La Figure ci-dessous montre le graphe crée par 20000 utilisateurs ayant employé le tag « sf ». Les deux pôles de ce graphe montrent à quel point l'analyse des réseaux d'utilisateurs peut aider à déterminer puis éventuellement enrichir la sémantique des tags, en les situant dans des ensembles d'utilisateurs.

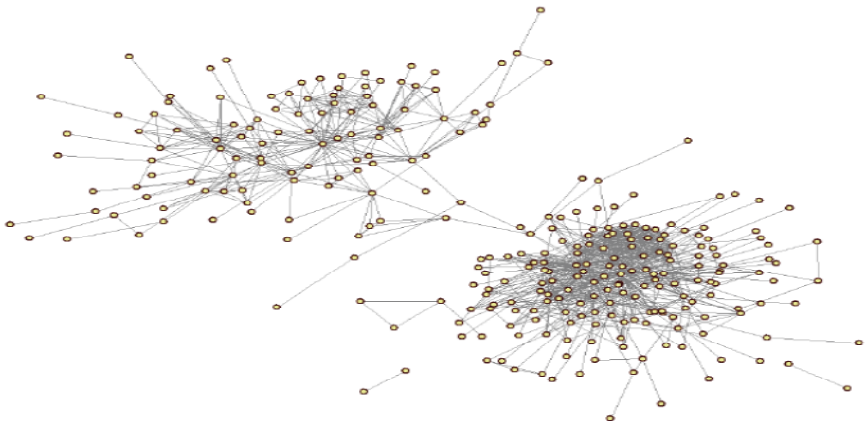

Figure 2.7. Graphe des utilisateurs de Delicious

La deuxième famille de travaux vise à modéliser les différents éléments des systèmes à base de tags (tags, actions de tagging ...) avec les technologies du web sémantique. Ces modèles, sont généralement des ontologies pour les folksonomies,ils permettent ainsi d'envisager les systèmes à base de tags comme partie intégrante du Web Sémantique, puisque ils sont représentés en RDF(S)/OWL. Ainsi [Gruber, 2007] propose un premier modèle étendant la notion tripartite classique et le redéfinit comme une relation faisant intervenir quatre éléments :

- *L'Objet.*
- *LeTag.*
- *L'utilisateur.*
- *La source.*

Le dernier élément (l*a source*) représente l'espace où est effectuée l'action de tagging exemple : Delicious. C'est cette dernière propriété qui enrichit la relation initiale et qui permet de distinguer deux actions de tagging d'un même utilisateur pour la même ressource mais sur deux espaces distincts. Gruber propose aussi d'ajouter des tags négatifs c'est à dire pour un utilisateur d'affirmer qu'un tag n'est PAS associé à une ressource. Ce procédé est utile pour éviter le phénomène de « mauvais » tags ajoutés par des spammeurs.

[Passant, 2009] quand à lui, y ajouta un autre paramètre supplémentaire au modèle tripartite classique, qui concerne non pas la source du tag mais carrément sa signification dans son contexte. Plus particulièrement, il distingua :

- La signification locale du tag, c'est-à-dire la signification particulière et non ambiguë d'un tag lors de l'action de tagging.
- Mais aussi, les significations globales du tag, qui peuvent lui-être associées si l'on considère le terme seul, hors contexte.

Il étend ainsi de la manière suivante le modèle de représentation tripartite d'une action de tagging en un modèle quadripartite où la signification (*Signification*) est ici considérée comme locale : *Tagging*(*Utilisateur,Ressource, Tag, Signification*).

On retrouve ainsi de nombreux outils qui se sont développés pour aider les internautes à relier leurs tags à des concepts d'ontologie, Une des applications les plus abouties est *Faviki*[23] permettant de tagguer des ressources mais seulement en employant le titre d'articles issus de Wikipédia. Il y a aussi le système MOAT de [Passant, 2009], Annotea, Semanlink qui se base sur des ontologies tels que la *angTag Ontology, SIOC*[24] *et SKOS.* On retrouve aussi GroupMe, qui propose aux utilisateurs de regrouper les tags par catégories pour faciliter la recherche d'information, représentant le tout avec sa propre ontologie.

Cependant, *nous trouvons personnellement* que ces approches cherchant à faire relier par les utilisateurs les tags à des concepts prédéfinis, demandent un certain temps et une énergie supplémentaire pour que l'utilisateur puisse choisir le bon concept, plutôt que de taper le premier terme lui venant à l'esprit comme c'est le cas actuellement. Il nous semble donc primordial de conserver les différents points de vue des utilisateurs, d'essayer de les coordonner mais sans qu'une vision commune ne soit obligatoire (comme c'est le cas avec les ontologies). Cette coordination des différents points de vue, reste pour le moment un problème peu abordé, ce qui pourrait bien entendu constituer plus tard une bonne perspective de recherche.

Á ce propos, on retrouve [Zacklad, 2007] qui pour sa part, suggère le recours à des ontologies semi-structures, appelés « ontologies sémiotiques » dans le cas où les différents points de vue des utilisateurs ne sont pas consensuels. Mais, ces ontologies sémiotiques nécessitent cependant le recours à un spécialiste au même titre que pour les ontologies classiques, et ne constituent donc pas un outil aussi spontané comme les folksonomies.

Il serait donc intéressant, d'utiliser un système combinant, à partir d'une analyse des différents points de vue des utilisateurs des communautés web, les traitements automatiques des approches citées précédemment et les contributions des utilisateurs via des interfaces ergonomiques. Ces traitements automatiques permettent comme nous l'avons précisé ci-

23. www.faviki.com
24. Semantically-Interlinked Online Communities

dessus, d'extraire des relations sémantiques entre tags en analysant la structure des folksonomies, et permettre à chaque utilisateur d'organiser les tags selon son propre point de vue, tout en bénéficiant des contributions de ses pairs.

Il serait aussi intéressant dans une deuxième étape, de fournir un moyen intégrant des fonctionnalités de structuration des tags, qui permettrais aussi de détecter les éventuels conflits entre points de vue des utilisateurs et de temporairement les solutionner par un agent automatique. Ce qui nous permettra d'exploiter ce point de vue cohérent pour enrichir chaque point de vue individuel avec les toutes autres contributions tout en garantissant une cohérence locale.

2.4. Le Web 2.0 à l'ère du mobile

Il est fort de constater à travers la section précédente, que le Web 2.0 en plus d'être une évolution d'un point de vue technologique, c'est aussi une évolution des usages du Web, parmi ces évolutions on retrouve la démocratisation du Web mobile et de l'aspect Web temps réel. La fusion de ces deux concepts est communément appelé *Web Squared*, pour marquer une phase transitoire entre le Web 2.0 et le Web du futur (omniprésent), C'est d'ailleurs l'initiateur du Web 2.0, Tim O'Reilly et un de ses collaborateurs qui ont décidés de lancé l'appellation ***Web Squared*** comme nouvelle génération Web, et pouvoir introduire et regrouper en son sein ces avancées technologiques du web mobile. Cette notion a donc été présentée lors du dernier *Web 2.0 summit*, le 25 juin 2009 et leurs a alors permis de définir dans un White Paper[25] les différents piliers qui composerai ce *Web²*. Nous avons ainsi pu à travers la lecture de ce White Paper, dégager les deux principaux enjeux du web squared :

- La standardisation des échanges entre bases de données pour croitre les chances d'interopérabilité des applications mobile sur le web.
- L'intensification des interactions entre capteurs GPS d'appareils mobile et l'environnement Web.

En effet, Tim O'Reilly et ses collaborateurs ont mis l'accent sur la notion de *Web of Things* qui selon eux, apportera une prise en compte des capteurs du monde réel dans le Web ; et précise ainsi que : *« les applications d'intelligence collective qui représentent un des principes clés du Web 2.0, ne sont plus seulement activées par des humains tapant sur des claviers, mais, de plus en plus, par des capteurs de nos téléphones et nos appareils photo par exemple, qui deviennent ainsi les yeux et les oreilles des applications. Ces capteurs de mouvement et de localisation permettrait d'indiquer où nous sommes, ce que nous regardons, et à quelle vitesse nous nous déplaçons. Ce qui permettrait de collecter des données, les présentées et les exploitées en temps réel, avec plus d'utilisateurs et de capteurs alimentant plus d'applications et de plates-formes.*

25. http://www.knowtex.com/posts/web-le-web-2-0-cinq-ans-plus-tard_3002

*En conséquence, les possibilités du Web ne croissent plus de manière arithmétique mais croissent de manière exponentielle. D'où notre thème pour cette année : **Le Web à la Puissance Deux**. »*

Ainsi, le Web² est donc présenté comme complémentaire au Web 2.0, une sorte de phase intermédiaire qui nous mènera vers de nouvelles évolutions du Web. Ce schéma ci-dessous (tiré à partir du site *http://www.blog-city.com/community/*) permet d'illustrer les notions fondamentales que pourrais englober le Web² :

Figure 2.8. Evolution du Web 1.0 au Web²

Parmi ces différentes notions qui constituent les fondements du web[2], on retrouve tout d'abord, les ***Implied Metadata*** ou Métadonnées implicites. Celles-ci sont générées automatiquement, à l'aide de nouveaux gadgets comme les appareils photo numériques capables d'associer une grande quantité de données aux photos tels que exemple les coordonnées GPS ou encore le modèle de l'appareil [cavanna, 2009]**.**

Toujours dans le cadre des *Implied Metadata*, on retrouve aussi, les systèmes auto-apprenants qui sont capables de générer des métadonnées à partir de renseignements issus de sources différentes, c'est le cas lors des recherches de photos numériques par exemple, et cela en ayant recours aux logiciels de reconnaissance faciale sur le web mobile, et c'est justement le principe mis en œuvre par Apple iPhoto.

On retrouve aussi de nombreux autres domaines d'application grâce aux données potentiellement fournies par les capteurs, comme c'est le cas des appareils domestique connectés au Web et qui pourraient contribuer à la surveillance en temps réel de la consommation d'énergie des ménages, pour économiser l'énergie. Il est prévu donc de faire communiquer les compteurs de gaz à Internet pour permettre aux fournisseurs d'énergie, de savoir exactement les habitudes de consommations de leurs clients et ainsi mieux anticiper les éventuelles recommandations d'optimisation de leurs facture.

Bien évidemment tout ceci est encore en train de se perfectionner pour pouvoir être exploité à l'avenir, et il semblerait qu'une autre version du web tendra elle aussi, d'exploiter tout ces concepts et c'est ce que l'on appelle aujourd'hui le Web Symbiotique.

Une autre notion du web[2] est celle d'***information Shadow***. Dans ce cas là, les Smartphones jouent un rôle crucial car ils permettent aux utilisateurs de transmettre des renseignements, d'images géolocalisées à des systèmes en ligne de n'importe quel endroit, et de ce fait, les gens pourront interroger « *The Data-On-The-Cloud »*[26] et ainsi permettre à ses utilisateurs d'avoir accès à un ensemble d'informations et de renseignements associées aux objets, aux personnes, à un lieu ou encore à un événement. Ce service existe déjà chez Google appelé *Lunettes de Google*[27]. Celui-ci permet à titre d'exemple, de prendre une photo d'un musée, qui sera ensuite transférée vers leur gigantesque base de données, le système analyse l'image, reconnaît les caractéristiques, il la localise sur une carte satellite, identifie la construction et fournit des informations sur son histoire et ses expositions en cours.

Une autre application d'*Information shadow* est celle des codes barres de *mobile tags* et *flash codes*, ceux ci permettent de prendre en photo un objet pour en récupérer des informations complémentaires dessus. Cependant, nous n'en croisons pas beaucoup d'application du genre pour le moment, mais parmi celles les plus fiables, on retrouve Microsoft Tag[28].

On retrouve aussi les applications de Réalité augmenté. Leur principe est simple, il suffit d'installer une application sur le téléphone et celle-ci récupère en temps réel des données pour les afficher sous forme de bulles qui sont autant de points d'intérêt sur le lieu ou l'on se trouve. L'application *Layar*, la plus aboutie actuellement ; divers services la pratique, notamment ceux liée au tourisme. Son champ d'application est vaste mais tout cela reste cependant en version expérimentale en attendant plus d'utilisateurs.

Et enfin, la dernière notion que nous aborderons dans cette partie est celle des *Data Ecosystems*, qui comprend les métadonnées de type déclaratives, cette notion représente un sous-projet du web sémantique au sein du W3C. Son principe est d'attacher à chaque mot important d'un texte une balise indiquant le contexte, et de structurer cette information. Cela est possible grâce à certaines couches de type RDF décrit les métadonnées et les annotations des données, le langage OWL décrit les descriptions et ajoute des contraintes logiques qualifiantes, et on obtiendra au final une collection ontologies, ceci permettrai de standardiser les échanges entre bases de données et permettant alors de croitre les chances d'interopérabilité entre applications, comme annoncé par Tim O'Reilly dans son livre blanc.

26. http://richard.cyganiak.de/2007/10/lod/
27. http://www.google.com/mobile/goggles/
28. http://tag.microsoft.com/home.aspx

2.5. Conclusion

Cet état de l'art nous a permis d'évaluer les principaux changements apparus au cours de la période Web 2.0. Nous avons aussi abordé deux aspects négatifs de celui-ci. En effet, le nombre d'utilisateurs sans cesse croissant et la simplicité de manipulation des outils, on aboutis à une surabondance d'informations, qu'il devient difficile à gérer et à organiser ; nous avons par la suite, exposé les principales solutions qui permettent de réguler cette masse d'informations, leurs faiblesses, mais aussi les recherches réalisées où en cours de réalisation.

Nous nous sommes aussi penchés sur les faiblesses liées au Tagging. Ce dernier est basé initialement sur la catégorisation des contenus par les utilisateurs eux-mêmes, ce qui leur laisse un libre choix dans l'indexation de leur contenu. Mais face aux conséquences chaotiques qu'ils engendrent, plusieurs travaux de recherches ont été menés ces dernières années dans le but de remédier à ces problématiques. Des travaux toutefois, fortement inspirés par les techniques du Web sémantique ; c'est d'ailleurs la raison pour laquelle nous pensons que la continuité du succès des Pratiques et Techniques 2.0, dépendra essentiellement de l'utilisation des technologies du Web Sémantique, ce qui ne peut être qu'un plus en termes de structuration et d'échange de données.

Cette fusion des approches Web 2.0 et Web sémantique, ouvrira certainement le champ à de nombreuses perspectives de recherches, visant à faire profiter les Outils 2.0 de l'utilisation de modèles communs pour représenter des métadonnées basés sur RDF, ceci se ferait grâce à l'utilisation d'ontologies permettant la structuration des connaissances produites via ces outils, en lieu et place des API actuellement hétérogènes. Où encore permettre l'utilisation de protocoles de requêtes et d'échange standardisés et de ce fait, réduire les insuffisances liées aux Web services actuels.

Enfin, pour conclure ce chapitre, nous avons présenté le « Web Squared », une initiative de Tim O'Reilly qu'il considère comme une continuité au Web 2.0 mais avec d'avantage de mobilité et de souplesse de navigation.

CHAPITRE 3 : VERS L'ÈRE DU WEB 3.0

Le précédent chapitre a permis de voir les différents apports en termes d'innovations techniques et fonctionnelles des outils Web2.0. Mais voila que maintenant, d'autres initiatives voient le jour. Ces dernières s'inscrivent dans la lignée de ce que l'on appelle : LE WEB 3.0. Toutefois, ce dernier prête déjà à controverse. En effet, il existe plusieurs directions qui se dessinent pour ce Web 3.0 (Web sémantique, Web des objets, Cloud Computing etc.), ce qui rend difficile de proposer une définition claire de celui-ci. Il est néanmoins certain que le Web sémantique représente l'un de ses piliers fondateurs.

Aussi, le Cloud Computing, représente une technologie permettant à ses utilisateurs de disposer de toutes leurs applications (Professionnel, Personnel…) sur des appareils à puissance réduite (PC, Mobile, Tablettes …), et de toutes leurs données à n'importe quel moment et n'importe quel endroit grâce au Cloud. C'est en quelque sorte tout un écosystème du Web qui est visé. Autant les applications Web mobile que les différentes architectures de Web services. C'est ce tout interconnecté qui crée ainsi le Web 3.0.

L'intérêt de ce chapitre est de passer en revue ces différents domaines à savoir : le Web sémantique, le Web de données et enfin le cloud computing, pour essayer de comprendre au mieux les avantages qu'ils pourraient apporter au Web de demain, ainsi que de diagnostiquer les principaux verrous scientifiques auxquels ils sont actuellement confrontés.

Première Partie : Le Web sémantique & de données

Parmi les différents articles de réflexions sur le Web 3.0, on retrouve une grande majorité dont [Mika, 2007] qui se limite à décrire celui-ci comme étant une période où l'on verra se développer des principes et technologies du Web sémantique ; certes cette vision n'est pas totalement fausse, mais elle est quelque peu faussée. En effet, il est important de rappeler que les principes et technologies du web sémantique n'ont pas attendu le Web 3.0 pour se développer. Voilà de nombreuses années que le RDF est exploité comme métalangage et que de nombreuses autres initiatives permettent de structurer l'information : pour la syndication, pour les formulaires ou encore les microformats.

Donc le Web 3.0 va au dela du simple Web sémantique, pour y ajouter la notion de Web de données (Linked Data), car malgré les efforts de standardisation des langages du Web Sémantique, il subsiste toujours un certain problème lié à la visibilité des données sémantisées, malgré que certaines ontologies ont permis d'entrevoir une démocratisation de ces données, le domaine est longtemps resté limité. C'est suite à ce constat qu'est née

l'initiative *Linking Open Data*[29] en 2007, avec l'objectif d'exposer en RDF un grand nombre de données déjà présentes sur le Web (dans des formats hétérogènes ou sous forme de simples documents HTML) et d'interconnecter celles-ci. Ainsi le Web des données parviendra à ajouter une vision plus pragmatique au Web sémantique, au sens où ce sont les données et les bases de connaissances qui seront mises en avant, et non pas seulement les ontologies et les possibilités qu'elles offrent.

Cette initiative a déjà permis de produire un nombre impressionnant de données liées, issues de différentes sources de données aussi diverses que DBpedia[30], Freebase[31] ou encore à partir des profils utilisateurs de Flickr ; tout ceci à l'aide de différentes stratégies utilisées pour produire ces liens entre données, en allant de la contribution manuelle des utilisateur [Hausenblas et al., 2008] à l'utilisation d'heuristiques plus poussées [Raimond et al., 2008], il reste cependant un certain problème de gestion d'ambiguïté qui se pose. Nous reviendrons sur cela au long de ce chapitre.

Toutefois, le passage au Web 3.0 ne suppose en aucun cas, à une refonte totale des usages du Web actuel, notamment ceux issus du Web 2.0. En effet, dans un premier temps ces deux visions (Web 2.0 et Web sémantique) peuvent paraître quelque peu divergentes du fait que le Web 2.0 soit plus centré sur les utilisateurs et que le Web sémantique soit plutôt centré sur les machines, mais nous pensons comme d'autres d'ailleurs [Gandon, 2006] [Gruber, 2008] [Ankolekar et al., 2008] que ces deux visions ne semblent pas si contradictoires, et que bien au contraire, elles peuvent et doivent chacune bénéficier des apports et travaux de l'autre communauté. Ceci permettra de converger vers un Web optimisé à la fois pour les humains et les machines, au niveau des modes de publication pour le premier et de la modélisation des données pour le second. C'est cette fusion qui, selon nous, permettra d'aboutir à un Web de Données issues d'interactions sociales tout en étant réutilisable de manière autonome via des agents logiciels. Il existe déjà plusieurs travaux dans ce sens, qui tendent à renforcer les outils et usages issus du Web 2.0 par une couche sémantique et qui permettra de réunir la puissance sémantique et la souplesse nécessaire à la participation des utilisateurs.

Ainsi nous nous intéresserons dans ce chapitre aux verrous scientifiques rencontrés par les technologies actuelles du Web sémantique, nous présenterons certaines solutions notamment en se basant sur des travaux récents où actuellement en cours dans le domaine. Nous reviendrons ensuite sur le concept de Web de données pour présenter le projet *Linking Open Data*.

3.1. Discussions et travaux récents autours du Web sémantique

Les travaux liés aux techniques, langages et pratiques visant à mettre en action les théories du Web sémantique sont nombreux, on constate d'ailleurs qu'ils commencent à fournir des résultats techniquement intéressants en termes de prototypes et d'architectures. Cependant, ces travaux souffrent pour la majorité d'entre eux de certaines problématiques

29. http://linkeddata.org/
30. http://dbpedia.org/About
31. www.freebase.com/

liées à leurs déploiements sur des systèmes aussi distribué que Internet.

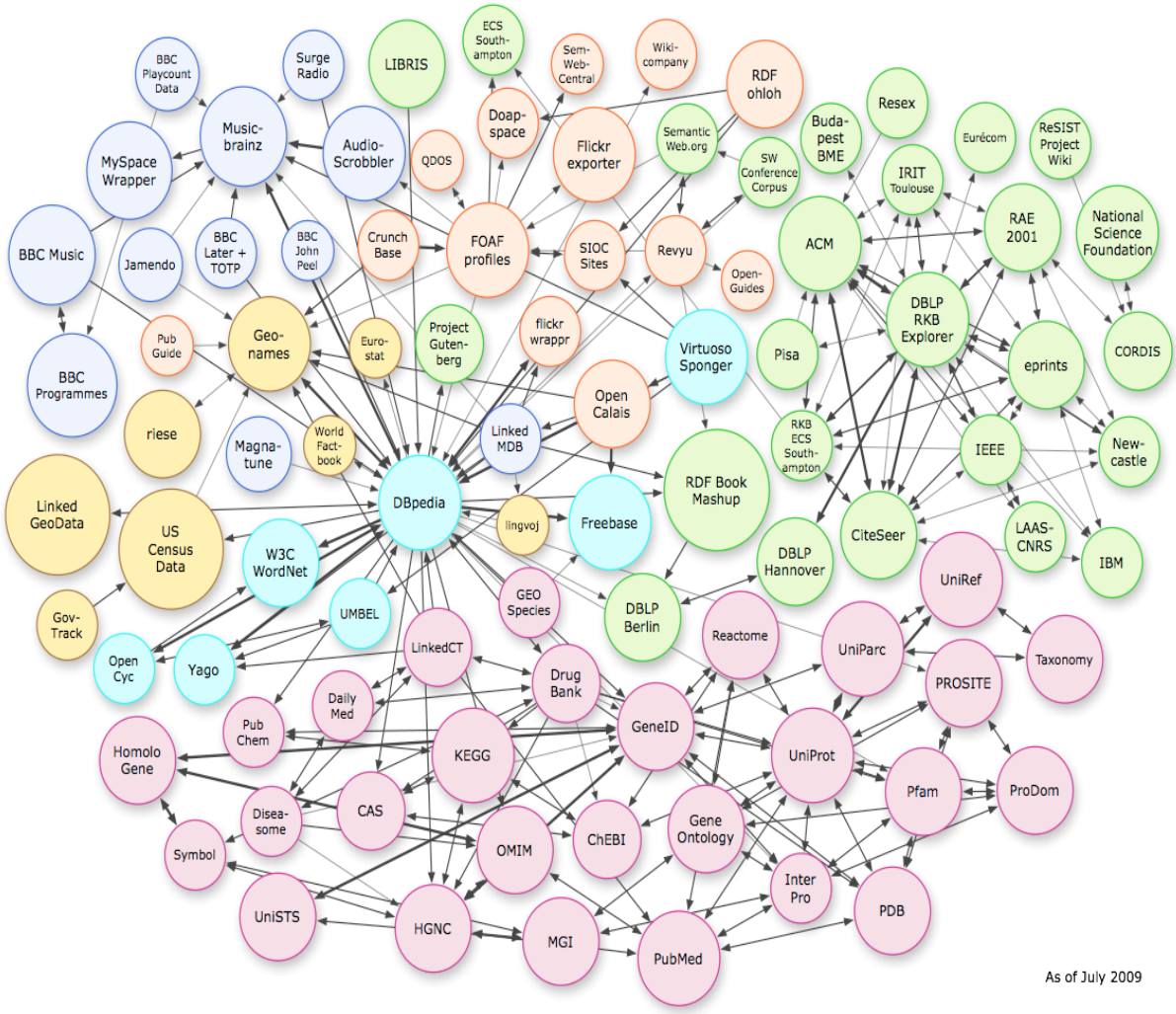

Figure 3.1. Diagramme « Linking Open Data cloud ».

Ceci dit, les choses commencent quelque peu à s'améliorer. D'ailleurs un premier changement constaté ces trois dernières années, est la médiatisation du Web sémantique autour de quelques produits et acteurs de référence. On retrouve Twine, Freebase, Powerset, Hakia, OpenCalais, DBPedia, le projet DataPortability ou encore l'API «Google Social Graph». Autre évolution majeure constatée, est l'arrivée à maturité des technologiques et plus précisément des langages du Web sémantique à savoir, RDFs, SPARQL et autres OWL, qui trouvent des terrains d'applications de plus en plus nombreux sur le web, ce qui permet de voir le nombre de données formalisées en RDF en constante augmentation. Cependant, l'évolution majeure du web sémantique dépendra toujours de la construction et de la prolifération des ontologies et de l'utilisation de métadonnées et annotation sémantique des ressources Web, c'est pour cette raison que nous essayerons de voir dans cette section les principaux verrous scientifiques liées à ces dernières.

3.1.1. Travaux autours de l'interopérabilité des Ontologies.

Les travaux et recherches autours des ontologies ont pris réellement leurs essor avec l'implication de chercheurs en représentation des connaissances, autour des langages et systèmes d'inférences notamment avec la définition du langage OWL (*Ontologies Web Language*) ; mais aussi avec les chercheurs de l'ingénierie des connaissances sur la question des méthodologies d'élaboration d'ontologies. Toutefois ces ontologies sont souvent trop spécialisés, hétérogènes et propre à des communautés bien déterminées, c'est ce qui a introduit une nouvelle problématique, celle de l'interopérabilité entre celles-ci. La question sera ainsi de savoir : « *Quel est actuellement le meilleur moyen, qui permet d'assurer au mieux, une interopérabilité cohérente entre ontologies de domaines hétérogènes ?* »

Parmi les certaines recherches consultés lors de l'élaboration de cet état de l'art. On retrouve entre autres [Djedidi, 2007], [Giunchiglia, 2008] et [Zargayouna & Nazarenko, 2010] dont les travaux portent sur la normalisation et l'interopérabilité des ontologies, leurs permettant d'être plus facilement partageables et réutilisables sur le Web.

Les travaux citées ci-dessus convergent tous vers une solution faisant intervenir le principe de « mapping » entre concepts des différentes ontologies, en effet les données sont annotées à l'aide d'annotations sémantiques qui décrivent leur signification, ensuite une correspondance entre données proches sémantiquement est creée, et des mappings servirons enfin à relier les ontologies qui manipulent ces données. Cependant malgré tous les efforts investis dans ce domaine, les résultats sont pour le moment au stade d'expérimentation et limités, car même si par exemple des méthodes dites sémantique considèrent la sémantique décrite pour deux ontologies plus ou moins proche, il se pourrait que ces ontologies soit incohérentes entre elles, certainement à cause de l'existence de différents points de vue pour la même donnée lors de son annotation sémantique, ce qui limiterai ainsi les résultats attendus.

Toutefois, D'après [Giunchiglia, 2008], l'interopérabilité sémantique est hautement dépendante du contexte et de la tâche, en effet toutes les données n'ont pas la même utilité pour chaque tâche, et que toute donnée n'est pas appropriée dans tout contexte. Pour réaliser l'interopérabilité sémantique, il est donc nécessaire de prendre en considération le contexte et la tâche, lorsque les mappings sont établis, évalués, ou utilisés. [Giunchiglia, 2008] proposa dans son article des suggestions intéressantes à ce propos, mais sans pour autant en formuler une solution qui pourrait bien être exploitée, d'un autre coté [Ben Abbés, 2010] donne une solution qui consiste à décomposer le problème de l'évaluation des ontologies en sous-problèmes suivant la nature des données manipulées ce qui pourrait bien servir à cerner le sens de l'ontologie pour ensuite poursuivre dans la même ligné que [Giunchiglia, 2008] et proposer par la suite une méthodologie pour déterminer ce qui fait partie du contexte, comment le collecter, le modéliser, et l'employer. *Ceci pourrait à notre avis*, constituer une bonne perspective de recherche pour la résolution de cette problématique.

3.1.2. Solutions à l'enrichissement d'Ontologies

Un autre important challenge à surmonter concernant les ontologies et de façon générale le Web sémantique, est celui de l'enrichissement d'ontologies. Ce processus peut être divisé en deux étapes à savoir : une phase *d'apprentissage ontologique* pour rechercher de nouveaux concepts et relations, et une phase de *placement de ces concepts et relations* au sein de l'ontologie, cette dernière étape est plus connue sous le terme de *Peuplement ontologique*. Nous nous intéresserons dans ce qui suit à ce concept de peuplement d'ontologies seulement, et tenterons d'entrevoir les solutions et travaux existants et qui nous permettrais de servir de repère à d'éventuelles recherches à ce propos.

Ainsi le peuplement d'ontologie, (action d'ajouter des instances à une base de connaissance), est pour le moment une tâche majoritairement assisté par les humains ou semi-automatisé via l'analyse de corpus de textes [Kiryakov et al., 2004] [Amardeilh et al., 2005]. En effet, après la découverte de termes candidats ou susceptibles à un changement, il est indispensable de détecter les relations entre ces nouveaux termes ainsi que celles qui les lies à l'ontologie initiale. Dans [Faatz *et al.*, 2002], les auteurs proposent pour cela une approche statistique basée sur la co-occurrence fréquente de termes candidats avec des termes de l'ontologie initiale. L'inconvénient majeur de ces travaux réside dans le fait qu'ils ne permettent pas l'ajout ; d'une manière précise ; des nouveaux concepts ni de relations entre eux dans une structure en constant changement tel est le cas avec le Web.

D'autres approches dans la littérature proposent d'utiliser des techniques de fouille de données [Jorio *et al.*, 2007], [Hernandez *et al.*, 2007]. Les travaux de [Neshatian *et al.*, 2004] se basent sur une méthode de classification afin de rapprocher les termes candidats contenus dans les textes des concepts présents dans l'ontologie. Le principe est similaire que celui explicité par [Giunchiglia, 2008] et que nous avons décrit dans la partie concernant l'interopérabilité des ontologies [Section 3.1.1.]. Cependant, ces solutions se basent pour la majorité d'entre elles sur des approches (semi-) automatique, généralement confiées à une équipe dédiée composée aussi bien d'experts du domaine que de spécialistes en ingénierie des connaissances, pour qu'ils puissent s'assurer de la qualité des données produites, à la fois en termes de valeur intellectuelle (via l'expert du domaine) et de qualité sémantique (via les spécialistes en ingénierie des connaissances). Au vu de l'importance du travail que devrais fournir l'équipe, les communautés Web semblent quelque peu se désintéresser de ces pratiques notamment, à cause des couts de maintien et d'évolution qui se révèlent trop élevés.

En effet, ces méthodes précédemment citées ont souvent fait leurs preuves au sein de réseaux restreins (Intranet), mais elles impliquent cependant l'impossibilité pour des contributeurs externes (issue du Web) de faire profiter l'équipe de leur expertise, puisqu'ils ne font pas partie du groupe destiné à maintenir ces Ontologies. Cela nous a conduits à la question suivante : *Y'a t'ils des solutions alternatives et plus automatisés, faisant intervenir d'avantage les pratiques collaboratives et participatives, pour pouvoir palier au manque d'enrichissement et d'évolution ontologique sur le Web ? Quel outil du Web collaboratif (Web 2.0) serait le mieux adapté à cet effet ?*

Première constatation, c'est qu'il n'existe pour le moment que très peu de solutions à proprement dites, exploitable à grand échelle sur le Web, grâce aux outils et pratiques issues du Web 2.0, toutes sont à l'état d'expérimentation. Nous pouvons citer à ce propos, le projet SIOC (*Semantically-Interlinked Online Communities*) qui propose de décrire les activités mais aussi les objets couramment utilisés sur les sites communautaires à savoir les blogs, les billets, les commentaires, ainsi que leurs relations. Il utilise des objets définis dans d'autres ontologies, comme FOAF[32] (pour décrire les personnes impliquées), SKOS, Dublin Core[33] et RSS (pour décrire les contenus), il se fonde ensuite sur le déploiement d'une base générique minimale de règles associatives entre items ne contenant que des associations entre termes non redondantes, afin d'enrichir son ontologie. L'avantage de cette solution est qu'elle se base déjà sur d'autres ontologies d'où l'aspect collaboratif, ce qui permet ainsi de valider et de s'assurer que les instances produites sont préalablement conformes à des modèles ontologiques.

Une autre initiative plutôt prometteuse pour l'enrichissement d'ontologies fait appelle aux systèmes Wikis, et plus précisément au Wiki sémantique. Ces derniers se définissent comme des wikis améliorés par l'utilisation des technologies du Web sémantique. Plus particulièrement, un wiki sémantique est similaire à ceux traditionnels, dans le sens où c'est un site Web dans lequel le contenu est ajouté par les utilisateurs. Ce contenu est organisé en pages éditables et indexables, accessibles à tous les utilisateurs. Cependant, contrairement au wikis traditionnel, ceux sémantiques ne se limite pas au texte en langage naturel. En effet, ils caractérisent les ressources et les liens entre celles-ci, et ces ressources sont ensuite formalisées et deviennent donc exploitables par une machine, à travers des processus de raisonnements artificiels.

L'avantage de ces Wiki réside dans le fait qu'il existe deux approches différentes pour leurs conceptions. La première, appelée *wikis for ontologies* ou *Wikitology*, qui est la plus utilisée, considère les pages comme des concepts et les liens typés comme des propriétés. Il se dessine ainsi une ontologie formalisée, dont la précision est le plus souvent renforcée par la catégorisation des concepts. Dans ce cas, la structure des données est généralement très souple et permet une grande liberté pour l'utilisateur mais ne garantit cependant pas l'utilisabilité des ontologies résultantes et cela alourdit la base de connaissances et nuit à son homogénéité.

La deuxième approche, *ontologies for wikis*, est celle qui nous intéresse le plus, car elle suppose généralement l'utilisation d'une ontologie déjà existante. Selon cette approche, le but du wiki consiste à fournir les outils permettant son peuplement par l'ajout d'instances et parfois de classes. Ces outils sont généralement des formulaires de choix multiples ou utilisant l'auto-complétion, et refusent par exemple la création de nouveaux types de liens ce qui garantie ainsi la cohérence de l'ontologie finale. Cette approche apparente le wiki à un éditeur de métadonnées, permettant de peupler l'ontologie.

32. www.foaf-project.org/
33. Schéma de métadonnées générique qui permet de décrire des ressources numériques

Au vu des avantages que procure cette deuxième approche, des travaux de recherches ont été menés dans ce sens notamment [Laublet, 2008] [Passant, 2008] et ont aboutis enfin à une implémentation concrète sous le nom de UfoWiki (*Unifying Forms and Ontologies in a Wiki)*, qui s'agit en quelque sorte d'un serveur de wiki, permettant à chaque utilisateur d'instancier une nouvelle instance pour sa communauté, le tout consultable via une interface unique.

La conception de ce service s'est fait sur la base d'une plateforme de Wiki existante, ce qui permet de bénéficier des développements relatifs à la partie wiki classique de l'outil (rétro-liens, historique des versions, etc.), et ainsi de ne pas troubler les utilisateurs en les confrontant à un nouvel outil.

Pour résumer *UfoWiki* repose sur les principes innovants suivants :

- Une représentation des connaissances basée sur des ontologies prédéfinies.
- Une interface utilisateur simplifiée pour le peuplement d'ontologies.
- Une utilisation immédiate des connaissances produites.
- Une réutilisation de données externes au Wiki.
- Des annotations mutualisées entre les différents wikis.

Toutefois quelques failles demeurent. En effet, du moment que *UfoWiki* repose sur l'approche *Ontologies for Wikis*, et que cette dernière ne garantit pas l'homogénéité de l'ontologie, puisqu'elle peut comporter plusieurs ressources ayant la même sémantique. Ainsi à l'heure actuelle, la seule solution semble être le contrôle manuel des ajouts par des ingénieurs des connaissances, cela peut en effet être pénible à l'échelle du Web mais reste tout de même le moyen le plus optimal en attendant de futures améliorations.

3.1.3. Annotation sémantique et hétérogénéité des ressources

Un autre axe de recherche du Web sémantique et un des objectifs à atteindre pour l'émergence du Web 3.0 est de maximiser la construction et l'utilisation des métadonnées permettant d'annoter sémantiquement les ressources, ceci afin d'améliorer la recherche d'information. Atteindre cet objectif dépend principalement du développement des ontologies, et des possibilités d'automatiser les méthodes d'annotation.

Techniquement, une annotation sémantique consiste à assigner à une entité (une chaîne de caractères, une phrase, un paragraphe, une partie de document ou un document) une métadonnée dont la sémantique est définie dans une ontologie. Elle a comme objectif d'annoter des concepts, des instances de concepts ou leurs relations. Cette annotation peut être stockée dans le document lui-même, ou dans un autre document référençant l'entité annotée par son URI. De nombreux projets existent dans ce sens, certains visent à construire des bases d'annotations comme DBpedia qui annote sémantiquement des pages wikipédia dans le langage RDF. Quand au projet *Linked Data* que nous verrons plus tard vise plus à décrire, à

partager et à connecter différentes ressources provenant de différentes sources telles que DBpedia, DBLP et Geo-names. D'autres projets ont pour objectif de faciliter l'annotation manuelle de ressources Web comme Annotea, un standard du W3C qui vise à améliorer la collaboration sur le web en annotant les documents dans le langage RDF.

Compte tenu du nombre de ressources disponibles sur Internet, l'annotation manuelle est une tâche longue et fastidieuse qui ne convient pas à l'échelle du Web. L'automatisation des techniques d'annotation est donc un facteur clé pour le Web 3.0 et son passage à l'échelle. Mais l'une des principales difficultés rencontrées réside dans le fait que les ressources Web sont souvent hétérogènes aussi bien du point de vue de leur format (HTML, pdf, gif, mpeg, avi,...) et de leur structuration que du point de vue du vocabulaire utilisé.

Pour limiter ces problèmes liés à l'annotation, la majorité des méthodes développées se focalisent souvent sur un domaine d'application bien particulier, sur une classe de sites Web ou encore sur des entités précises comme l'annotation des noms de personnes, des adresses e-mail, ou des pays. Mais Etablir des annotations grâce aux méthodes précédentes pose une double difficulté.

- ***La première difficulté*** concerne la segmentation du texte parce qu'il est souvent difficile d'identifier précisément les éléments textuels à annoter. C'est un problème connu en reconnaissance d'entités, en effet : *Est-ce que le déterminant et le nom classifieur font partie de l'entité ou est-ce que seul le nom doit être annoté ?* D'une manière générale, certaines connaissances ontologiques ne se traduisent pas par un simple mot ou expression. C'est souvent un fragment large ou une phrase complète qui véhicule l'information. Le fait de prendre ou non en compte ces annotations dépend largement de l'objectif visé.

- L'ambigüité inhérente à la langue soulève ***une seconde difficulté***. Car lorsque le Fragment textuel est ambigu, il faut choisir l'élément ontologique avec lequel il doit être mis en correspondance. Le processus d'annotation suppose alors une étape de désambigüisation qui repose généralement sur des indices figurant dans le contexte du fragment à annoter.

Pour effectuer ces annotations sémantiques en résolvant ce problème de segmentation et de désambigüisation, la communauté du Web sémantique se penche à des solutions permettant soit de représenter conjointement lexiques ou terminologies et ontologies, soit étendre l'ontologie avec des connaissances permettant de mettre en correspondance le texte à annoter et l'ontologie avec laquelle il est interprété. T*els est le défi actuel concernant les annotations sémantique au sein du Web sémantique.*

Parmi les travaux recensé dans ce sens, on retrouve, [Montiel-Ponsoda et al., 2007], [W3C, 2009], [Buitelaar et al., 2006] et [Cimiano et al., 2007], toutefois ces travaux supposent la nécessité d'un niveau lexical ou terminologique de base, entre la ressource ontologique et le texte. Ce niveau lexical pose deux problèmes importants, le premier est celui de sa représentation et le second celui de sa constitution. Ainsi, nous pensons qu'il est plus intéressant d'étendre les ontologies par des règles d'annotation sémantique qui permettrai de

tirer le meilleur parti des outils de TAL[34] (ceux qui produisent chacun à leurs niveau des annotations linguistiques). Elle a également le mérite de distinguer clairement le processus d'analyse et la tache d'interprétation.

À ce propos, les travaux de [Adeline Nazarenko et al. 2009] proposent d'étendre l'ontologie par des règles d'annotation qui s'expriment sous la forme de patrons, Ces travaux présente selon nous de nombreux intérêts. Tout d'abord, le pouvoir expressif des patrons est très grand. Il va de la simple représentation d'une liste de mots à des expressions complexes basées sur des annotations de hauts niveaux (entités, termes) et autorise l'expression de règles de flexion ou de désambigüisation.

Ensuite, les patrons sont compréhensibles et modifiables par une personne tout en étant interprétables, voir calculables, par un ordinateur, ce qui permet d'envisager leur acquisition automatique à partir d'un corpus annoté. Enfin, ils peuvent s'exprimer dans de nombreux formalismes largement connus voire standardisés comme les expressions régulières. Cependant ces travaux restent pour le moment au stade d'expérimentation, et appellent d'ailleurs à un double prolongement. En effet, il faudra tout de même tester l'approche proposée en intégrant un module d'annotation sémantique dans une chaine d'annotation existante, chose qui n'est pour le moment pas fait. Et enfin, la question de l'acquisition des règles est également une piste à explorer car il est d'autant plus important de pouvoir apprendre (semi-)automatiquement les règles qu'elles varient d'une ontologie à l'autre.

3.2. Web des données : « Projet Linking Open Data »

Présente dans la feuille de route du Web sémantique écrite par Tim Berners-Lee, l'expression « *Web of data* » qu'on traduit par « Web de données » n'a été vraiment utilisée qu'à partir de 2006 suite à la parution de la note « *Linked Data* » du même auteur. Cette note est d'une importance fondamentale d'abord pour le Web 3.0 et pour le Web en général, puisqu'elle reprend les principes de bases du Web sémantique à savoir, celui d'établir des liens entre les données exposées et distribuées sur le Web.

Ainsi, elle constitue le point de départ pour une nouvelle expension du Web sémantique notamment avec le projet du W3C « Linking Open Data » visant à placer sur le Web des données structurées en RDF et à offrir des cas d'utilisations réels et simples à l'aide des technologies du Web sémantique. En novembre 2009, le Web de données était constitué de 13,1 milliards de triplets répartis au sein de différents ensembles de données couvrant les domaines aussi diverses que les données multimédia, les données du Web social, les données géographiques et statistiques, les données bibliographiques...

[Tim Berners-Lee, 2006] à édicter quatre principes formalisant les pratiques courantes à la base du Web de données, il a ainsi précisé qu'à l'aide de ce dernier on sera ammener à :

34. Traitement Automatique des Langues.

- Utiliser des URI pour identifier les choses ;
- Utiliser des URI accessibles via HTTP ;
- Ouvrir aux machines l'accès aux données en utilisant les standards RDF et SPARQL ;
- Exprimer l'URI des objets liés.

Il a aussi indiqué que plus les données seront étroitement liées à d'autres données, plus leurs valeurs et leurs utilités augmentent. Donc, on voit bien à travers ces quatre principes que, Tim Berners cherche non pas à remettre en cause le principe fondateur du Web, mais au contraire s'appuie sur son architecture de base à savoir, le système des URI et le protocole HTTP, pour ainsi en faire à juste titre une extension de celui-ci. Mais en lieu et place des documents, il s'agit tout simplement de lier des données faisant partie de deux ensembles de données distincts d'où l'expression « hyperdata » parfois employée.

Toutefois, effectuer ces liens s'avère complexe car la nature de ces derniers n'est pas très riche pour le moment, et que dans la majorité des cas l'équivalence d'identité entre deux ressources se fait simplement avec des propriétés qui ne sont même pas définies dans le vocabulaire des langages du Web sémantique. En contre parti, les Sept milliards de triplets que représenterait le Web actuellement (Source W3C), on compte seulement 142 millions de liens entre les ensembles de données. Ce qui est insignifiant devant le nombre de données disponibles actuellement sur le Web. Ainsi face à ce constat, l'enjeu majeur pour le Web de données (et par conséquent pour le Web 3.0) est de ***Voir croitre le nombre de données formalisées en RDF pour aussi voir en paralléle croitre celui des données liées.*** Ceci nécessite donc un très gros effort de représentation de données en RDF.

3.2.1. Représentation des données avec RDF(S)

Bien entendu, cette problématique de représentation des connaissances avec RDF(s) ne date pas d'aujourd'hui ; différentes sérialisations permettent déjà de représenter des assertions modélisées en RDF. C'est le cas de N3 [Berners-Lee, 2006], Turtle [Beckett et Berners-Lee, 2008], ou encore les annotations sémantiques vues plus haut, celles-ci semblent d'ailleurs l'une des solutions à avoir donné le plus de résultat. Mais cela dit les annotations semblent pour le moment limitées. En effet, les annotations sémantiques sont en général représentées sous la forme de documents RDF indépendants des éventuels documents HTML /(X)HTML associés, par conséquent il est facile de remarquer que cela introduit généralement un problème de duplicité d'informations. Ainsi l'ajout de métadonnées directement au sein de pages Web est aujourd'hui une des voies prometteuse à cette problèmatique, cette proposition est déjà au coeur de différents travaux comme c'est le cas avec eRDF ou RDFa [Adida et Birbeck, 2008] qui permettent l'inclusion directe d'annotations RDF au sein de documents (X)HTML.

Dans cette même optique d'annotations intégrées au sein même des pages, nous pouvons également citer les microformats[35], effort communautaire qui offre aussi la

35. *www.*microformats.org

possibilité de définir certaines données structurées (événements, contacts ...) au sein de pages Web via de simples attributs de balises semblables à RDF. Ceux-ci ne sont malheureusement pas aussi puissants que RDF(S) en termes d'expressivité de plus, ils ne bénéficient pas d'une bonne ouverture à l'évolution comme c'est le cas pour les ontologies par exemple, puisqu'un microformat ne peut évoluer qu'après consensus de la communauté. Mais ils sont néanmoins utilisés plus fréquemment sur le Web malgré ses différentes faiblesses.

Le projet GRDDL pour (*Gleaning Resource Descriptions from Dialects of Languages*) est aussi une autre alternative ; celui-ci est un mécanisme mis au point par le *W3C* à partir de 2009. Il sert à extraire les informations présentes sur une page Web au format *XHTML***,** de traduire les différents dialectes XML transformant ainsi les annotations RDFa en données RDF brutes, qui peuvent être ensuite utilisées comme n'importe quelles données RDF natives. Ceci permetterai ainsi de faire le pont entre ces différentes visions cités ci-dessus.

Mais malgré tous ces efforts recentie dans la représentation et standardisation de données formalisés, il faut reconnaître qu'elles sont toujours peu nombreuses, d'autant plus que parmi ce nombre impressionnant de données disponibles et notamment celles formalisés en RDF, on retrouve qu'une très grande majorité d'entre elles appartiennent à des institutions privés. De ce fait, les données qui en ressortent sont aussi de type propriètaire et donc contraire au libre accés comme convenue sur le Web. C'est devant ce constat que des voix se sont élevées pour promouvoir une circulation plus libre de ces données. Par conséquent, la préoccupation de la communauté du Web de données est actuellement de sensibiliser et encourager d'avantage ces producteurs de données à *: Se joindre au mouvement d'ouverture des données et d'un libre partage de celle-ci, ce qui serait tout aussi complémetaire au projet Linked Data*.

Nous nous interesserons dans ce qui suit à la compréhension de ce mouvement d'ouverture des données, pour mieux saisir l'intérêt de ce projet. Nous présenterons brièvement les initiatives qui tentent de structurer le paysage des données ouvertes, nous décriverons ensuite quelques acteurs déjà investis dans cette nouvelle vision du Web.

3.2.2. De quelles données parle-t-on ?

Selon Wikipédia : « Les données ouvertes sont à la fois une philosophie et une pratique, qui exige que certaines données soient mises à disposition de chacun librement, sans restrictions liées à des droits d'auteurs ou tout autre mécanisme de contrôle »[36].

On recense d'un autre coté différentes prises de position des personnalités du Web, qui témoignent de leurs points de vue à travers différents évenements. Ainsi côté américain, nous avons analysé essentiellement les discours de :

36. http://fr.wikipedia.org/wiki/Open_data

- Tim Berners-Lee, lors de son intervention au TED[37].
- Tim O'Reilly et John Battelle à travers leur texte de référence « Web Squared: Web 2.0 Five Years On ».
- Ainsi que les articles des principaux éditorialistes des publications en ligne de référence comme ReadWriteWeb[38].

Nous nous référons aussi à l'article[39] de Frédéric Cavazza. Et les points principaux qui ressortent de leurs discours à propos de ces données ouvertes sont :

- ✓ Une conviction forte que les données, quoi qu'il arrive, vont s'ouvrir (O'Reilly et Battelle).
- ✓ Une volonté de convaincre que les données doivent s'ouvrir au plus vite (Tim Berners Lee, Alexander Korth).
- ✓ Les données réelles et les données du web sont amenées à se croiser (O'Reilly et Battelle).
- ✓ Les données constituent une ressource économique forte : « Les entreprises doivent apprendre à exploiter des données temps réel comme des signaux essentiels qui alimentent une boucle de rétroaction beaucoup plus efficace pour le développement de produits, le service à la clientèle, et l'allocation des ressources » (O'Reilly et Battelle).
- ✓ Les données ouvertes permettent de « fabriquer des applications qui créent de la valeur et des opportunités au Web sémantique » (Vivek Kundra).
- ✓ Les données ouvertes sont la base d'une intelligence collective croissante (O'Reilly et Battelle).
- ✓ Les données libérées et reliées sont indispensables à la recherche scientifique, à l'innovation et doivent permettre de faire face aux grands défis de l'humanité (Tim Berners Lee).
- ✓ Elles permettent de rendre la gestion du monde meilleure (Tim Bernars Lee).

On remarque cependant à travers ce survol, se pointer différentes confusions où approximations autour de la nature des données ouvertes, cela nous a amené à se poser la question suivante : ***(i) « De quel type de données parle-t-on réellement ? »***. A cet effet, nous avons distingué deux grandes catégories de données, cette classification est notamment en fonction des producteurs de ces données, on retrouve ainsi les données produites par des entreprises privées (site de e-commerce) ou publiques (données cartographiques) et les données produites par des acteurs publics (gouvernements, collectivités locales, universités, centres de recherche, instituts statistiques etc). Mais cette hiérarchisation semble toutefois contestée pour être encore plus étendue. En effet, les données produites par des utilisateurs courant du web (d'une façon générale ceux du Web 2.0) semblent mises à l'écart. Ce qui nous conduis au final à se demander, à qui appartiennent véritablement les recommandations, les commentaires, les notations, les réseaux d'amis, les interactions entre utilisateurs ?. Si les conditions générales d'utilisation tranchent souvent en faveur d'une propriété du site qui les

37. www.**ted**.com/talks/**tim**_berners_lee_on_the_next_web.html (2009)
38. fr.readwriteweb.com/
39. Article de Cavazza « du contenu roi aux données reines »

héberge, ceci semble ainsi contraire à cette démarche d'*Open Data* (Données Ouverte).

A ce propos la figure 3.1 fait apparaître sous plusieurs couleurs les différents types de données présentes actuellement dans le Web de données, et notamment celles qui participent que ça soit à ce mouvement d'ouverture mais aussi à celui du Linked Data. Cela nous permet en quelque sorte de réponde à notre question initial *(i)*. Ci-dessous une description détaillée du diagramme :

On retrouve ainsi ***des ressources d'intérêt général*** (en bleu clair) qui recouvrent essentiellement les données issues de dictionnaires ou d'encyclopédies. De ce point de vue, le projet le plus emblématique est *Dbpedia*. Initiative lancée en 2007, *Dbpedia* vise à extraire les informations structurées de Wikipedia et à rendre cette information disponible avec les technologies du Web sémantique. Pour ce faire, Dbpedia s'appuie sur les « infobox », un encartage généralement présent à droite d'un article de la Wikipedia constituant une « carte d'identité » de la ressource décrite, les liens reliant les différentes versions de la Wikipedia, les catégories, les liens présents dans l'article... Mis au point et maintenu par Universität Leipzig, Freie Universität Berlin et par différentes sociétés commerciales, ce projet met à disposition 274 millions de triplets RDF sur 213 000 personnes, 328 000 lieux, 57 000 albums musicaux...

Les ressources issues du « web social » (en saumon) recouvrent les projets de conversion des Web Services de sites Web 2.0 aux technologies du Web sémantique, l'exposition des données personnelles en utilisant le vocabulaire FOAF ou l'exposition des sites Web Sociaux (forums, blog, wikis...) avec le vocabulaire SIOC précédemment cité.

Les ressources géographiques et statistiques (en jaune) recouvrent les projets d'exposition de données géographiques et les projets de mise à disposition des données publiques dont une bonne partie sont des données statistiques. Parmi les projets de mise disposition des ressources géographiques, on peut citer *Geonames*, système d'information géographique sous licence libre, qui référence et donne les coordonnées géographiques de 8 millions d'emplacements où encore *LinkedGeoData* qui est au service *OpenStreeMap* ce que Dbpedia est à Wikipedia et qui contient 320 millions de points géoréférencés et 25 millions d'itinéraires.

Les ressources multimédia (en bleu foncé) recouvrent des conversions de bases de données musicales en ligne comme Music Brainz ou Jamendo, mais aussi des initiatives plus originales comme celles de la BBC, cherchant à valoriser et à mettre à disposition dans une logique d'ouverture les données accumulées depuis de nombreuses années. L'originalité de la démarche réside dans la réutilisation de données existantes dans le Linked Data enrichies de leurs propres données pour construire des sites Web conviviaux à destination des utilisateurs et manipulables par les machines. De ce point de vue, le site BBC Music constitue une réussite et un exemple précurseur pour la mise à disposition de données culturelles.

Les ressources médicales et biologiques (en violet) recouvrent tous les ensembles de données qui ont été agrégés par le projet *Bio2RDF* et le groupe d'intérêt « *Semantic Web*

Health Care and Life Sciences » du W3C. En effet, le modèle de graphes constitue le modèle de référence pour échanger les données biologiques réparties sur le réseau, leur mise à disposition peut être cruciale pour accélérer la recherche dans la découverte de remèdes contre les maladies. Avec cet ensemble, le domaine de la biologie médicale démontre tout l'intérêt scientifique que revêt l'accès ouvert aux données brutes de la recherche.

Les données bibliographiques (en vert) recouvrent à la fois les projets de catalogue de bibliothèques, des bibliographies sélectives type DBLP (bibliographie en informatique) ou *Semantic web dog food* (bibliographie de différentes conférences dans le domaine du Web sémantique) que des conversions selon les principes du Linked Data de Web services existants comme c'est le cas avec *RDF Book Mashup.*

Deuxième Partie : Le Cloud Computing

3.3. Le Cloud Computing, une autre tendance du Web 3.0

Le Cloud Computing, ou informatique dans les nuages, est un paradigme assez récent. Il permet d'assurer une certaine mobilité et omniprésence du Web et qui représente un des piliers du Web 3.0. Sa réelle mise en application a pris place au début des années 2000 et à émergée à la suite du Web 2.0 notamment grâce à Google et Yahoo. Le Cloud consiste ainsi en une communication entre un serveur frontal et un ensemble de machines virtuelles qui hébergent une ou plusieurs applications. Ainsi, le visiteur a accès à des applications dont l'exécution ne dépend pas uniquement d'un seul serveur web, mais sans toutefois influer sur son temps de réponse.

La définition du *Cloud Computing* de Wikipedia souligne que : *« L'informatique dans les nuages (en anglais, Cloud Computing) est un concept majeur faisant référence à l'utilisation de la mémoire et des capacités de calcul des ordinateurs et des serveurs répartis dans le monde entier et liés par un réseau, tel Internet »*.

Le Cloud (Figure 3.2) permet donc de fournir un ensemble d'applications puissantes sans utiliser la mémoire, la puissance de calcul et la capacité de stockage d'un seul serveur. Le visiteur se connecte sur le site du Fournisseur des services de Cloud, utilise les applications qui lui sont proposées sans avoir conscience qu'il accède à des machines différentes (virtuelles ou non). Les données éventuellement ainsi produites sont elles aussi stocker sur des serveurs distants. Le Cloud comprend généralement un où plusieurs des concepts ci-dessous :

a. L'IaaS (Infrastructure as a Service)

Il s'agit de la mise à disposition, à la demande, de ressources d'infrastructures dont la plus grande partie est localisée à distance dans des centres de données. Les serveurs, postes de travail, et imprimantes peuvent être facturés en fonction de leur utilisation. Le client loue par exemple de la CPU, de la mémoire pour le stockage de données et le coût est directement lié

au taux d'occupation. Une analogie peut être faîte avec le mode d'utilisation des industries des commodités (électricité, eau, gaz) ou des Télécommunications.

b. Le PaaS (Platform as a Service)

Il s'agit de la mise à disposition pour une entreprise d'environnements techniques pour développer des applications qui fonctionneront à distance comme pour le SaaS. Mais l'IaaS inclut des outils de personnalisation et une intégration à l'existant ou à d'autres programmes hébergés. L'objectif étant de proposer un environnement modulaire capable de combiner plusieurs fonctions et processus métier, voire plusieurs technologies en provenance de divers éditeurs.

c. Le SaaS (Software-as-a-Service)

Il s'agit de la mise à disposition d'un logiciel, non pas sous la forme d'un produit que le client installe en interne sur ses serveurs, mais en tant qu'application accessible à distance comme un service, par le biais du Web. Les clients ne payent pas pour posséder le logiciel en lui-même mais plutôt pour l'utiliser. Ils l'utilisent soit directement via l'interface disponible, soit via des API fournies (souvent réalisées grâce aux *web services* où à l'architecture REST).

La contrepartie avec ces types de service est que le client n'a pas directement accès à ses données. Il dépend donc totalement du fournisseur et doit lui faire entièrement confiance pour ce qui est de leurs confidentialités et de leurs sauvegardes.

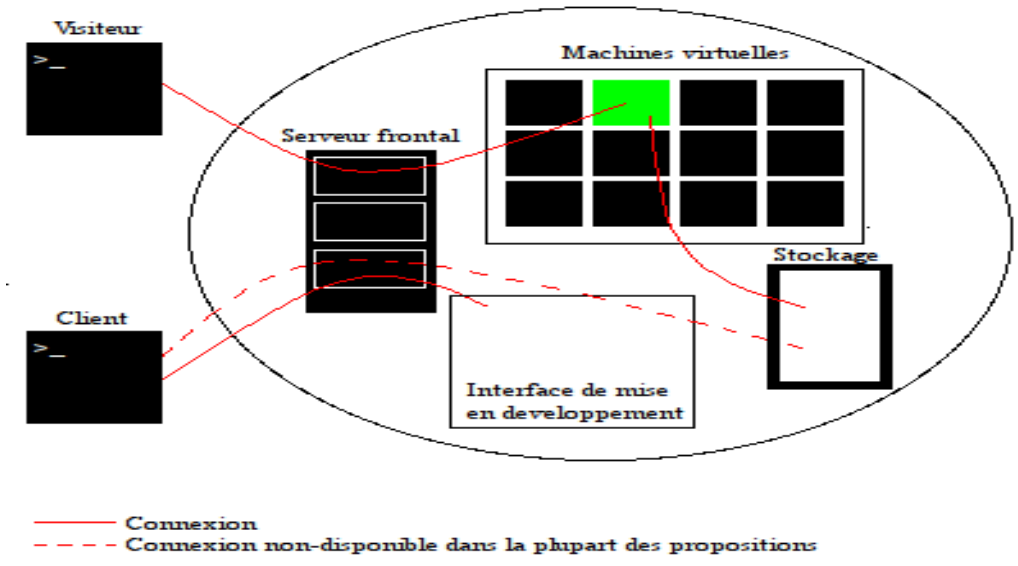

Figure 3.2. Fonctionnement du Cloud Computing.

3.3.1. Intérêt du Cloud

De la même façon que la virtualisation[40], un système de Cloud permet une grande évolutivité. On peut facilement et sans danger pour les applications déjà disponibles ; de rajouter des machines au Cloud pour une plus grande réactivité ou pour fournir des applications supplémentaires. De plus, s'il est constitué de machines virtuelles (ce qui est généralement le cas), le Cloud permet une réduction réelle des coûts pour l'appropriation de solutions Web. De plus, les ressources utilisées sont mieux rentabilisées notamment pour les entreprises. Aussi, si une application présente une faille, seul le système qui l'accueille pourra être mis en danger, et de ce fait toutes les autres applications ainsi que la machine frontale sont protégées. Les données et les applications étant hébergées et souvent sauvegardées sur des machines distantes, on peut y accéder de manière permanente et de n'importe quel endroit et être assuré de leur pérennité.

3.3.2. Mode de fonctionnement typique

Un Cloud comprend généralement les éléments suivants :

- ***Proxy http*** : Point d'entrée des demandes, gère le SSL[41].
- ***Cache http*** : Permet de répondre plus rapidement à une requête en plaçant une partie du contenu dans un cache.
- ***Serveur frontal*** : Gère les requêtes en lançant des machines virtuelles adéquates ou en communiquant avec des machines virtuelles adéquates déjà lancées.
- ***Machines virtuelles*** : Ensemble de serveurs (serveurs Ruby avec framework Web par exemple), accueillant chacune une application. Elles doivent pouvoir être lancées rapidement et indépendamment pour répondre le mieux possibles aux demandes des visiteurs.
- ***Base de données SQL ou système de stockage*** : Base de données pour chaque application, avec duplication et téléchargement pour le client (la base de donnée peut être externe au Cloud), système de stockage présentant les mêmes avantages.
- ***Cache mémoire*** : Cache mémoire pour les applications web permettant un accès rapide (par exemple à des fragments de pages).

Chaque machine virtuelle accueille un environnement spécifique au langage utilisé par le client. Une application utilise une ou plusieurs machines virtuelles suivant sa complexité. Le système de fichier peut être en lecture seule puisque (les données sont stockées ailleurs, il s'agira alors seulement de pouvoir exécuter l'application présente sur la machine virtuelle). L'environnement doit accueillir un serveur d'applications lui aussi spécifique au langage utilisé qui accueillera le serveur web. Et enfin, le serveur accueille l'application du client.

40. Consiste à faire fonctionner sur un seul ordinateur plusieurs systèmes d'exploitation
41. Secure Sockets Layer. Protocole de sécurisation des échanges sur Internet

Dans un souci de généricité, on peut exiger que l'ensemble des applications soient stockés sur un serveur de stockage et que la machine virtuelle y accède au moment de l'exécution. Ainsi, on peut créer des images de machines virtuelles typiques, valables pour un grand nombre d'applications, ce qui les rendent plus légères d'un point de vue taille.

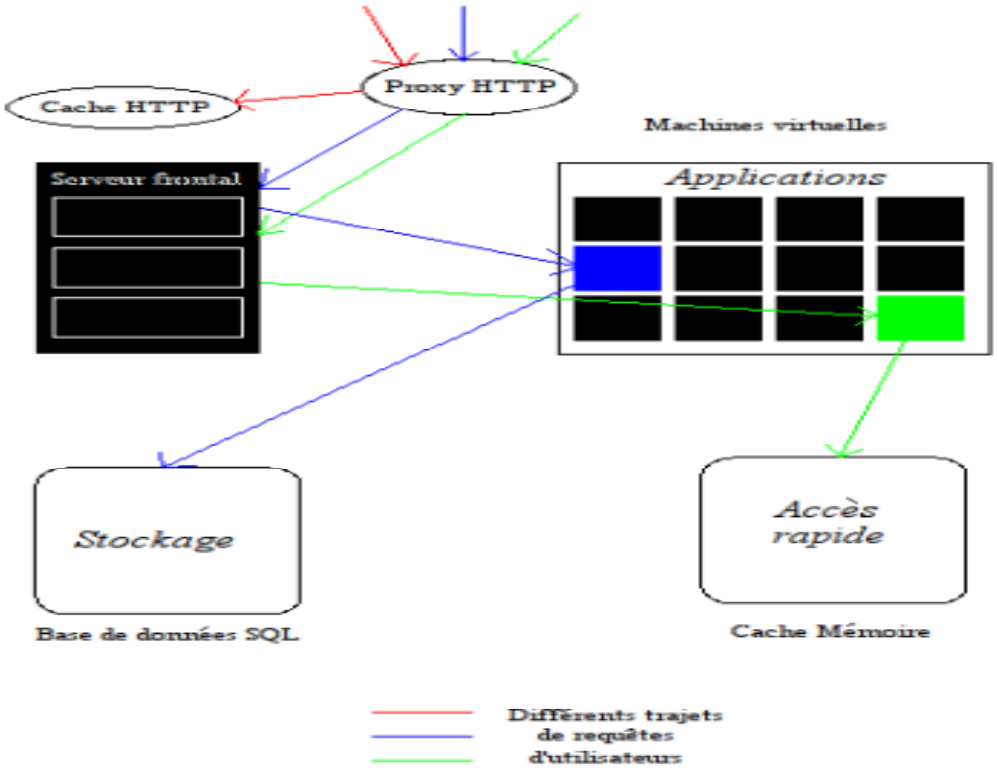

Figure 3.3. Fonctionnement détaillé

3.3.3. Usages du Cloud dans le Web 3.0

Les usages d'un Cloud dépendent principalement du point de vue adopté. Un Cloud se destine d'une part à un simple internaute, une entreprise ou autre (gouvernement, association, etc.). L'émergence du Web 3.0 dépendra essentiellement du succés du SaaS (Section 3.3, page 55). En effet, pour profiter pleinement de la mobilité et l'accessibilité du Web 3.0, il faut au préalable permettre aux internautes d'accéder aux services fournis via le *SaaS* quel que soit l'ordinateur qu'il utilise (plus besoin d'installer des logiciels équivalents à ceux proposé sur chacun de ses ordinateurs), et sans soucis de versions car l'application fonctionne de manière identique sous GNU/Linux, Mac, Windows ou tout autre système disposant d'un explorateur internet (iPhone, Android, Blackberry, PDA, ...). Garantissant ainsi une forte mobilité et une indépendance totale du dispositif utilisé.

Aussi, le mouvement des données ouverte pourrait bel est bien lui aussi profiter de cette opportunité. En effet, il faut donner les moyens techniques d'utiliser et de lire ces données, ce qui implique en principe d'importants efforts liées au déploiement

d'infrastructures ou de logiciels et par conséquent des dépenses de maintenance. Le *Cloud Computing* apparaît ici donc comme une solution souple, complémentaire pour s'affranchir des contraintes organisationnelles liées à la réutilisation de données ouvertes sur le Web.

Synthétiquement, on peut retenir donc quatre apports positifs du *Cloud Computing*:

1. *Une meilleure flexibilité technique.*
2. *Une continuité de l'activité à tout moment et n'importe quel endroit.*
3. *Une mise en œuvre aisé des concepts du Web des données et sémantique.*
4. *Une montée en charge selon la demande (réelle).*

3.3.4. Problématiques du Cloud Computing

Il est clair que le *Cloud Computing* n'est pas sans risques, puisqu'il entrainerait des exigences accrues en matière de fiabilité, de sécurité, de confidentialité. En effet, en externalisant un système d'information, on perd de ce fait la confiance de la fiabilité et sécurité des données. A cela s'ajoute des problèmes de standards et la non interopérabilité des Clouds, un point qu'il ne faut surtout pas négliger.

3.3.4.1. La sécurité dans le Cloud

Pour mieux comprendre les risques réels que pourrait apporter le Cloud Computing en termes d'atteintes à la sécurité des données, on se doit d'abord de prendre en compte l'une des principales caractéristiques de celui-ci. En effet, la notion de virtualisation et notamment les problèmes auxquels elle est confronté se répercutent systématiquement sur le bon fonctionnement du Cloud Computing, parmi ces problématique, on retrouve celles liées à la *Multi-tenancy*[42], soit le fait que plusieurs tenants/internautes exécutent simultanément la même instance d'une application ou processus métier déployé.

En effet, C'est ce principe même qui permet une réutilisation de fragments de processus métiers par les différents tenants et c'est cette réutilisation qui contribuera à optimiser le temps de service. Néanmoins, elle influe négativement sur la sécurité et l'authenticité des données utilisées, car le fait de réutiliser des fragments déjà calculés par d'autres tenants, peut amener des processus métiers malicieux déployés par des tenants malveillants, de vouloir récupérer des données sensibles seulement en réutilisant ces fragments. Ainsi, l'enjeu majeur ici est de pouvoir proposer des solutions pour la préservation des données sensibles lors de leurs externalisation pour les Clouds de type *SaaS*.

A ce propos, certains travaux de recherche notamment [Noman Mohammed, 2009], [Thomas Trojer, 2009] se penchent sur des techniques permettant la sécurisation des Mashup de bases de données. Tandis que d'autres eurent pour rôle la mise en œuvre d'algorithmes d'anonymisations [Latanya Sweeney, 2002]. Ces recherches ont cependant accès

42. Architecture logicielle, où une seule instance du service Web s'exécute sur un serveur.

sur les méthodes de chiffrement de données, ce qui permet en effet de remédier au problème de confidentialité des données, mais qui néglige toutefois celui de la réutilisation de fragments cité ci-dessus. Ainsi en guise de perspective, il serait tout à fait intéressant de pouvoir proposer des méthodes formelles, d'anonymisation des fragments de processus, afin de protéger la vie privée des tenants lors de la réutilisations de ces derniers, pour enfin pouvoir les implémenter sur des plateformes SaaS opérationnelles.

Il est aussi nécessaire de mettre en place des contrôles précis avant toute production de Cloud, ainsi que renforcer les sessions sécurisées via VPN, SSL où https, le cryptage des données en transits, les vérifications de l'intégration des données en E/S, favoriser le couplage du Cloud avec les différents profils et politiques d'accès de l'internaute. Où même encore la mise en place de techniques de monitoring et d'outils spécifiques pour chaque utilisateur, ce que les offres actuelle ne permettent d'avoir.

3.3.4.2. La question de l'interopérabilité

Les offres de Cloud (SaaS, PaaS et IaaS) ne manquent pas, citons entre autres Amazon Web Services EC2, Ikoula Cloud Iaas, Google App Engine, Cisco, IBM Cloud, Microsoft Windows Azure ou encore Sogeti etc. Aujourd'hui, la plupart de ces offres reposent sur des architectures assez propriétaires qui peuvent rendre les migrations difficiles en cas de problème sur l'un, où d'un simple changement de fournisseur. Ainsi, la question de l'interopérabilité entre services Cloud devient de plus en plus préoccupante.

Il existe de nombreux freins à cette interopérabilité, parmi lesquels on retrouve le manque d'API. En effet, il faut préciser que certaines applications *SaaS* ne disposent pas d'API ; dans ce cas l'utilisateur sera ainsi fortement dépendant de son fournisseur. En revanche quand celles-ci existent, elles sont soit au format REST (REpresentational State Transfer) ou SOAP (Simple Object Access Protocol). On voit ainsi que l'existence d'API, et leur complétude fonctionnelle, représente donc un vrai enjeu à l'émergence du Cloud.

Nous concluons ainsi que, l'interopérabilité ne va pas de soi et il faudra bien choisir son *PaaS* et *IaaS* tant que l'on ne disposera pas de SDK, API où de spécifications ouvertes, reconnues. Toutefois, des accords entre les différents acteurs du Cloud sur une éventuelle interopérabilité des différentes solutions via des standards sont donc à l'heure actuelle une bonne solution, en attendant mieux.

3.4. Conclusion

Ce chapitre nous a permis de présenter trois différents concepts essentiels pour la compréhension du Web 3.0. On a d'abord discutés de certains enjeux nécessaires à une meilleure intégration des techniques Web sémantique au sein des services Web. Dans la seconde section nous avons présenté l'initiative *Linking Open Data*. Celle-ci correspond, comme nous venons de le voir, à une large ouverture de données qui était autre fois exploités que part ses détenteurs, cette ouverture permettrai de les exposer à un large public

d'utilisateurs pour en créer, grâce aux outils du Web sémantique, de nouveaux services Web à valeur ajoutés. Et enfin dans la dernière section nous avons vu à titre d'information ; l'intérêt de la technologie *Cloud Computing* pour le futur Web, et pu distinguer deux des principales lacunes qui pourraient ainsi freiner son développement.

CHAPITRE 4 : LE WEB 4.0 EN PERSPECTIVE

Web ubiquitaire, Web omniprésent ou encore Web des Objets… tous ces termes font référencent à un monde de connexions encore plus denses, entre les hommes et les objets. Des connexions permanentes de plus en plus invisibles et omniprésentes. C'est cette vision que l'on appelle aujourd'hui LE WEB 4.0. Ainsi, dans ce chapitre, après avoir défini le terme Web 4.0, nous tenterons dans la deuxième section d'évaluer son stade de développement et d'énumérer les solutions nécessaires à son émergence. Et pour finir en guise de perspective nous nous interrogerons sur les conditions et éléments qui conditionnent son développement.

4.1. Définition du WEB 4.0

Entre les capteurs d'environnement, les puces RFID, les solutions de nommage ou encore les solutions de gestion de l'information de contexte, le Web 4.0 est composé de nombreux éléments complémentaires ayant chacun leurs propres spécificités, et par conséquent sa définition semble elle aussi perplexe. En effet, à la lecture des articles[44] sur le sujet, on remarque que leurs auteurs sont dans la majorité des cas tous d'accord pour dire que le Web 4.0 est du moins un Web omniprésent, et qu'il n'existe toujours pas de définition standard, unifiée et claire de celui-ci. Certaines définitions insistent sur ses aspects techniques, tandis que d'autres se concentrent plutôt sur les usages et les fonctionnalités. Ainsi, pour mieux appréhender la notion, nous proposons d'abord les deux principaux éclaircissements ou orientations que nous retrouvons souvent dans les ouvrages, pour ensuite exprimer dans une seule définition, ce que représente ce Web 4.0 tout en restant dans le contexte le plus clair et cohérent possible.

- **Conceptuellement : L'apparition d'identités nouvelles pour les objets.**

Certains définissent le Web 4.0 comme « des objets ayant des identités et des personnalités virtuelles, opérant dans des espaces Web intelligents et utilisant des interfaces intelligentes pour se connecter et communiquer au sein de contextes d'usages variés »[43]

D'autres émettent l'hypothèse que celui-ci représente une révolution permettant de connecter les gens et les objets n'importe où, n'importe quand, par n'importe qui. Ces définitions, mettent en effet l'accent sur la dimension ubiquitaire du Web, mais personnifient les objets en leur attribuant une intelligence et capacité de communiquer. Cependant, elles ne reflètent pas encore la dimension concrète liée aux usages de celui-ci.

43. Internet of Things in 2020. Roadmap for the Future. 2008
44. http://www.itu.int/itunews/manager/display.asp?lang=fr&year=2005&issue=09&ipage=things&ext=html

- **Techniquement : Une extension de nommage et convergence des identifiants**

Techniquement, le Web 4.0 est une extension au système de nommage actuel du Web et traduit une convergence des identifiants numériques au sens où il sera possible d'identifier de manière unifiée des éléments d'information numérique et des éléments physiques (comme un appareil électronique) tout cela d'une manière directe et instinctive grâce à l'utilisation de systèmes d'identifications électroniques (puces RFID, processeur et communication Bluetooth etc.). Ainsi il n'y aura pas besoin de saisir manuellement le code de l'objet, le réseau s'étend jusqu'à lui et permettra ainsi de créer une forme de passerelle entre les mondes physique et virtuel.

Définition :

> *« Le WEB 4.0 est un réseau de réseaux qui permet, via des systèmes d'identification électronique normalisés et unifiés, et des dispositifs mobiles sans fil, d'identifier directement et sans ambiguïté des entités numériques et des objets physiques et ainsi de pouvoir récupérer, stocker, transférer et traiter, sans discontinuité entre les mondes physiques et virtuels, les données s'y rattachant ».*

4.2. Composantes essentielles à l'émergence du Web 4.0

Le Web 4.0 ne se résume certainement pas à une technologie spécifique. Il désigne plutôt diverses solutions techniques (RFID, TCP/IP, technologies mobiles etc.) qui permettent d'identifier des objets, de capter, stocker, traiter, et transférer des données dans les environnements. Actuellement, l'enjeu majeur n'est pas tant d'inventer de nouvelles technologies mais de simplement perfectionner celles qui existent déjà, de les connecter, et de les intégrer. Nous listons ci-dessous les principales classes de solutions nécessaires au fonctionnement du Web 4.0 :

4.2.1. Les solutions RFID - Radio Frequency Identification systems -

Les solutions RFID font partie de la catégorie des technologies d'identification automatique. Elles sont en général utilisées pour fournir une identité électronique à un objet inanimé ou animé. Le sigle RFID recouvre un ensemble de technologies et d'applications très variées qui dépendent de paramètres tels que la portée, la bande de fréquence utilisée, le prix, l'encombrement, ou encore la consommation d'énergie, mais dépendent aussi de marqueurs/capteurs, lecteurs, et de logiciels pour traiter les informations collectées.

4.2.2. Les environnements d'intelligence ambiante

Les environnements d'intelligence ambiante servent d'interface entre les applications Web et les utilisateurs ou capteurs RFID. Ils ont un rôle critique dans les solutions RFID car ils permettent de gérer l'interface entre ces différents systèmes. Dans le cas des solutions RFID, ils assurent l'extraction des données RFID depuis les lecteurs, puis filtrent ces données, les agrègent et les transmettent après distillation.

Mais d'une façon générale, ces applications doivent prendre en compte le contexte dans lequel les utilisateurs évoluent (le lieu, la position sociale ou hiérarchique ou l'activité par exemple) pour adapter leur comportement. Ces applications doivent donc ensuite pouvoir prendre en compte dynamiquement l'arrivée de nouveaux éléments dans l'environnement (utilisateurs ou dispositifs de capteurs), les informations de contexte en provenance de l'environnement et enfin parvenir aux applications entrantes ; ces flux d'informations ne peuvent pas être déterminés à l'avance et doivent se construire pendant l'exécution. Les modèles de gestion de l'information de contexte existants pour le moment ne traitent pas ou peu cet aspect dynamique de l'informatique diffuse.

4.2.3. La standardisation EPCglobal

À ce jour, si plusieurs plateformes existent pour permettre de créer de nouvelles formes de réseau qui fait le lien entre le physique et le virtuel (Wi max, Bluetooth …), une seule solution semble pour l'instant suffisamment normalisée, adapté et reconnue pour pouvoir être utilisée à l'échelle du Web. L'architecture du réseau EPCglobal fut conçue par l'Auto-ID Center[45] et développée par la suite par l'EPCglobal.

La standardisation EPCglobal permet de détecter les puces RFID ayant une identification unique non ambiguë, nommée *Electronic Product Code* (EPC). Par la suite, les autres données relatives aux objets sont stockées et accessibles via le Web. Le réseau EPCglobal dispose par ailleurs de *l'Object Naming Service* (ONS) et de *l'EPCInformation Service* (EPCIS). L'ONS attribue une référence à l'information de l'objet sauvegardé via le réseau EPC, ce qui permet de retrouver les objets RFID à travers le réseau (son fonctionnement est calqué sur le *Domain Name System* utilisé pour internet).

EPCIS offre quant à lui une interface pour accéder aux données RFID mémorisées, et échanger ces données RFID entre le réseau EPCglobal et les environnements d'intelligence ambiante. L'avantage du réseau EPCglobal est de permettre l'utilisation de puces peu onéreuses et une architecture au périmètre ajustable. Il est important de noter que pour tracer les objets RFID dans le réseau EPCglobal, une des solutions proposées (Discovery Service) demande à ce que le mouvement de ces objets soit publié de manière continue auprès d'un où plusieurs serveurs référence de mises à jour.

On voit ainsi que le Web 4.0, notamment grâce à EPCglobal permettra progressivement d'élargir la notion de réseaux, en construisant un autre réseau de capteurs pour des objets, et contribuera également à structurer des types de réseaux inédits en tissant entre objets et individus des formes de structuration aussi nouvelles que celles déjà présentes sur le Web 2.0. Cependant, pour mieux comprendre ces évolutions, il serait donc essentiel de ne pas limiter l'analyse à la dimension technique du phénomène mais de considérer ses composantes sociales ainsi que les possibles interactions entre ses utilisateurs.

45. www.autoidlabs.org/

4.3. Principaux enjeux de développement pour le WEB 4.0

Même si de nombreux progrès ont été réalisés dans le cadre de la standardisation mais aussi des solutions RFID, il reste encore beaucoup de points à résoudre. Dans cette section, nous listerons les principaux éléments à prendre en compte dans les futurs travaux de développements du Web 4.0. Ainsi, deux enjeux majeurs se posent actuellement, l'un est d'améliorer la performance des solutions actuelles en garantissant une meilleure interopérabilité et un niveau de sécurité optimal. Et enfin garantir l'interopérabilité des standards utilisés.

4.3.1. Améliorer la performance des solutions.

Pour bien comprendre les problématiques actuelles, il faut distinguer deux classes d'amélioration possible : *Premièrement*, améliorer l'interopérabilité des solutions ubiquitaires, permettant ainsi de faire communiquer celles-ci entre elles, *Deuxièmement*, renforcer la sécurité à la fois pour les données mais aussi pour les personnes utilisant les solutions RFID.

4.3.1.1. Garantir l'interopérabilité.

Le succès du Web 4.0 dépendra essentiellement de l'interopérabilité des solutions d'intelligence ambiante. En effet, il est essentiel que tous les composants des solutions soient interopérables. Il faudra par exemple, que les lecteurs RFID soient compatibles avec la majorité des puces utilisées. Mais aussi, mettre en œuvre des solutions qui s'adapteront à des contextes hétérogènes pour garantir un fonctionnement fiable tout au long de l'utilisation.
La question de l'interopérabilité se pose également quand on envisage la convergence entre les solutions RFID et les solutions mobiles. Enfin, la solution à cette problématique dépend pour le moment essentiellement du standard EPCglobal. Ainsi les propositions des entreprises de télécommunication seront bien évidemment déterminantes, mais pour le moment aucune solution viable n'est adoptée, car la question de l'interopérabilité ne se résumerait pas seulement aux simples solutions techniques.

4.3.1.2. Garantir la sécurité.

Il faudrait aussi mettre en œuvre un système qui assure une sécurisation des données. En effet, si une solution RFID est mise en œuvre entre un client et un fournisseur, il faut que les deux parties partagent un certain nombre d'informations en commun mais dans un même temps, des données utiles pour le fournisseur ne doivent pas nécessairement être transmises au client. A titre d'exemple, *Joel de Rousnay*[46] précise que dans le secteur de la santé, ce type de problème est évidemment très sensible car les informations détenues sur chacun des patients ne doivent pas être accessibles par toutes les parties du secteur.

46. Conseiller de la présidence de la Cité des Sciences à Paris & Prospectiviste

4.3.2. Les enjeux de standardisation

Les standards actuellement candidat au Web 4.0 reprennent dans l'ensemble les même principes du Web actuel (DNS, TCP/IP, etc.) et des systèmes de code-barres. Ainsi le protocole TCP/IP et le langage XML restent toujours des références pour le Web 4.0. L'ONS est un système issu du DNS qui assure le nommage d'internet. De la même façon, les logiques de l'*Electronic Product Code* dérivent de celles utilisées pour le système du code-barres. Cependant, cette continuité des standards pourrait probablement limiter techniquement certains de ses développements et applications. Par ailleurs, si des milliards de données supplémentaires devaient transiter par le réseau, la capacité du Web à y faire face sera mise en cause pour des raisons physiques liées aux infrastructures(en raison notamment de la capacité des routeurs).

Pour répondre à ces nouveaux défis de sécurité et de fiabilité liés au Web 4.0, certains pensent que des mesures incrémentales peuvent suffire. Le passage de l'IPv4 à l'IPv6 (*Internet Protocol* version 6) est une illustration de cette démarche d'évolution. Ce nouveau protocole doit permettre l'augmentation de 2^{32} à 2^{128} du nombre d'adresses disponibles mais aussi de renforcer la sécurité, et enfin de mieux gérer l'arrivée massive de données selon un protocole particulier. Il agit sur la fragmentation des paquets de données pour les rendre ainsi plus « manipulable » par les routeurs. Il commence d'ailleurs a être proposé par les FAI aux clients professionnels, mais son déploiement tarde, faute d'une demande massive de nouvelles adresses IP. Pourtant de nombreuses applications existent et plusieurs *patchs* sont déjà disponibles pour une multitude d'applications[47], qui restent toutefois peu exploitées. Mais au-delà de la question de l'élaboration de standards, il est aussi fondamental de penser à leur interopérabilité, et ce d'autant plus que le WEB 4.0 est un réseau nécessitant un niveau d'interopérabilité élevé vu la diversité de ses composants. Actuellement, certaines avancées notables sont apparues, à l'image de NFC (*Near Field Communication*). Les solutions NFC permettent, via un système sans fil, de connecter deux dispositifs électroniques à courte distance, toutefois, ce type de standardisation est limité pour une utilisation à grande échelle.

47. http://www.ipv6.org/v6-apps.html

CONCLUSION GÉNÉRALE

Dans ce présent état de l'art, L'objectif était de prendre conscience des divers changements de paradigmes qui s'opèrent actuellement au sein du Web, en mettant en avant certaines caractéristiques de ces derniers, ainsi que l'apport de chacune d'eux pour l'écosystème Web. Pour cela nous nous sommes basés sur une classification hiérarchique à la « Web x.0 », nous permettant ainsi de distinguer les spécificités propres à chaque version.

On s'est par ailleurs, intéressé aux lacunes habituellement rencontrées sur le Web, ce qui nous a principalement amené à se pencher sur les techniques et pratiques de recherche d'informations pertinentes. En effet, face à la gigantesque quantité de données hétérogènes disponible sur le net, les outils et services de recherche d'information sont sans cesse remis en cause. A ce propos, nous avons abordé le manque de justesse des moteurs de recherche basés sur l'indexation par mot clés, du fait que leur fonctionnement interne intègre peu les approches centrées sur la ré-indexation collaborative où celles basées sur les processus adaptatifs. Les techniques de personnalisation sont elles aussi de mise, car en plus d'être des moyens efficaces au filtrage informationnel, c'est aussi des solutions combinant qualité, pertinence, et forte confiance des résultats retournés, comme le montre le modèle « Web Of Trust » de [Massa et al, 2004] (Chapitre 2, Section 2.2.2), ce qui représente un plus en terme de recherche d'information.

Nous avons aussi mis en évidence, les solutions mettant en œuvre les principes et langages du Web sémantique, que se soit en utilisant RDF et les diverses ontologies de domaines comme c'est souvent le cas, ou carrément en étendant le modèle de représentation tripartite, qui est à la base du langage RDF, selon les besoins du service à effectuer (Chapitre 2, Section 2.2.3.). Ces solutions nous ont aussi permis d'entrevoir certaines lacunes notamment liées à l'interopérabilité des ontologies mais aussi aux annotations sémantiques, et qui se pourrait être pénalisantes pour les éventuelles applications du Web sémantique, voir même à celles déjà mises en œuvre. Cela dit, à travers ce petit tour d'horizon, nous avons certainement pris conscience de l'intérêt que pourrait apporter l'approche du Web sémantique aux différentes outils et pratiques issues du Web 2.0. En effet, la sémantisation de ces outils, à l'image des Wikis sémantique où encore des MashUp sémantique, permettra d'outre passer les problèmes d'hétérogénéité des protocoles utilisés, mais aussi des formats de données ainsi produits. Cela influera positivement sur l'exploitation des données qui seront ainsi présentes sur le Web.

PERSPECTIVES DE RECHERCHE

À l'issu de ce mémoire, différentes perspectives de recherche s'offrent à nous. En effet, il serait certainement intéressant de proposer :

- Une approche basée sur le contexte (dans la continuité des travaux exposés à la Section 3.1.1. § 3) pour améliorer l'interopérabilité entre ontologies hétérogènes et évoluant de manière autonome. L'hypothèse sur laquelle cette perspective est basée est que les applications ou services du Web devraient avoir des ontologies qui représentent les concepts correspondant aux données qu'elles manipulent, pour qu'elles puisses être réconciliées avec d'autres ontologies d'applications voisines, sans que cela influe sur la qualité et la facilité de leurs échanges.

- Mettre en place une méthode automatique pour l'annotation sémantique de données multimédia au sein des plateformes de partages. En effet, cela serait certainement faisable en utilisant OSIRIS, qui est un outil d'annotation et de recherche sémantique de ressources multimédia fondées sur les graphes conceptuels. Le plus serait ainsi de permettre d'utiliser les Tags, largement répandus sur ce type de plateformes, pour l'enrichissement des ontologies qui cohabitent au sein de sein de l'outil OSIRIS.

- La Mise en place de procédés avancés permettant l'exploitation de données RDF de plus en plus nombreuses sur le Web, notamment via le projet *Linking Open Data*. Plus particulièrement, il nous semble intéressant de réfléchir à la manière dont celles ci peuvent être utilisées avec pertinence en termes de navigation, recommandation, réutilisation et découverte d'information. Il nous parait également intéressant d'y intégrer à nouveau un aspect social pour identifier des communautés d'intérêt ou des réseaux s'établissant autour de ces données.

- Mettre en place une infrastructure qui permettrait l'échange des informations de contexte, suffisamment générale pour être utilisée par différents services Web d'intelligence ambiante, et suffisamment spécifique pour couvrir les informations fournies par les dispositifs déjà utilisés comme les services de localisation, et suffisamment flexible pour accepter et exploiter dynamiquement l'introduction de nouveaux dispositifs.

TABLE DES FIGURES

LISTE DES TABLEAUX

BIBLIOGRAPHIE

[Adeline Nazarenko et al. 2009] A. Nazarenko, N. Hernandez, N. Nada, E. Sardet, *Apport des outils de TAL à la construction d'ontologies,* Acte de conférence : Lors de la Conférence d'Ingénierie des connaissances, Hammamet, Tunisie , Vol. tome 2, Mai 2009.

[Adida et Birbeck, 2008] B. Adida, M. Birbeck, RDFa Primer 1.0. In Proceedings of Working Group Note 2008, World Wide Web Consortium. San Francisco, USA. October 2008. http://www.w3.org/TR/xhtml-rdfa-primer/.

[Adomavicius et al., 2005] G. Adomavicius and A. Tuzhilin. Toward the next generation of recommender systems : A survey of the state-of-the-art and possible extensions, Acte de Conférence : Conference of IEEE Transactions on Knowledge and Data Engineering, Volume 17, Number 1, University of Tier, Germany, 2005.

[Amardeilh, 2005] F. Amardeilh. Web Sémantique et Informatique Linguistique :propositions méthodologiques et réalisation d'une plateforme logicielle. Thèse Doctorat de l'Université de Nanterre - Paris, France, Mai 2005.

[Ankolekar et al., 2008] A. Ankolekar, M. Krötzsch, D. Thanh, D. Vrandecic. The Two Cultures : Mashing up Web 2.0 and the SemanticWeb. Journal of Web Semantics, Volume n°05, 2007, Leibniz.

[Aymen *et al.,* 2009] M. Aymen, F. Krief, Négociation de niveau de service dans les environnements ubiquitaires, Acte de conférence : 9éme Conférence Internationale sur les Nouvelles technologies de la répartition, 2009. Montréal. CANADA.

[Beckett et Berners-Lee, 2008] David Beckett et Tim Berners-Lee, Turtle - Terse RDF Triple Language. In Proceedings of Team Submission, World Wide Web Consortium, CERN, January 2008.

[Ben Abbés, 2010] S. Ben abbés, Evaluation de classes sémantiques pour la construction d'ontologies, Acte de Conférence : 21èmes Journées Francophones d'Ingénierie des Connaissances, Nîmes, France. 2010.

[Berners-Lee et al., 2001] Berners-Lee, T., Hendler, J., and Lassila, *The semantic Web. (Version française),* Article de Revue : Scientific American, May 2001, p. 29-37.

[Berners-Lee et al., 2006] T. Berners-Lee, Y. Chen, L. Chilton, D. Connolly, R. Dhanaraj, J. Hollenbach, A. Lerer, D. Sheets. Tabulator : Exploring and Analyzing linked data on the Semantic Web. Acte de Conférence : In Proceedings of the 3rd International Semantic Web User Interaction Workshop, Athens, Georgia, USA. 2006.

[Berners Lee, 2008], Tim Berners Lee, Résumé de la Conférence « Linked Data Planet », New York, 2008.

[Buitelaar *et al.,* 2006] P. Buitelaar, M. Sintek, M. Kiesel, *A multilingual/multimedia lexicon model for ontologies*. Acte de confèrence : In ESWC, Budva, Montenegro, June 2006.

[Calvier, 2010] F. Calvier, Découverte de Mappings dans un Système Pair à Pair Sémantique : Application à SomeRDFS, Thèse de Doctorat de l'Université Paris-Sud 11, Octobre 2010.

[Candillier et al., 2007] L. Candillier, F. Meyer, M. Boullé. Etat de l'art sur les systèmes de filtrage collaboratif. International Conference on Machine Learning and Data Mining MLDM Germany, July 2007.

[Cao, 2006] T. Cao, Exploitation du web sémantique pour la veille technologique, Thèse de Doctorat en Sciences de l'université de Nice-Sophia Antipolis, 2006.

[Cavazza, 2010] F. Cavazza, *Du contenu roi aux données reines*, Post du 19 juillet 2010 de www.fredCavazza.net, 2010.
http://www.fredcavazza.net/2010/07/19/du-contenu-roi-aux-donnees-reines/

[Chaari, 2007] T. Chaari, Adaptation d'applications pervasives dans des environnements multi-contextes, Thèse de doctorat de L'institut national des sciences appliquées de Lyon, 2007.

[Cimiano *et al.*, 2007] P. Cimiano, P. Haase, M. Herold, M. Mantel, *Lexonto : A model for ontology lexicons for ontology-based nlp*. Acte de confèrence : In Proceedings of OntoLex - From Text to Knowledge, (Workshop at the International Semantic Web Conference), Busan, South Korea, 2007.

[Claypool *et al.*, 1999] M.Claypool, M. Gokhale, A. Miranda. Combining content-based and collaborative filters in an online newspaper, *Proceedings of ACM SIGIR Workshop on Recommender Systems*, Berkeley, Californie, 1999. Page 15

[Cliquet, 2010] G. Cliquet, *Méthode d'innovation à l'ère du Web 2.0*, Thèse de doctorat de l'école nationale supérieure d'arts et métiers (Spécialité Génie Industriel), 2010. Paris.

[Djedidi, 2007] R. Djedidi, Medical Domain Ontology Construction : a Basis for Medical Decision Support, Acte de Conférence : Proceedings of the 20th IEEE International Symposium on Computer-Based Medical Systems (CBMS'07), Special track on Healthcare Knowledge Management, Maribor, Slovenia, June 2007.

[Emonet, 2009] R. Emonet, *Description Sémantique de Services et d'usines à services pour l'Intelligence Ambiante*, Thèse de Doctorat du Groupe Grenoble INP (Spécialité Informatique), 2009. Grenoble. France.

[Ereteo *et al.*, 2009] G. Ereteo, *Analyse des réseaux sociaux et web sémantique: un état de l'art*, In Processing of Social management shared knowledge representations, ISICIL. 2009.

[Ereteo, 2011] G. Ereteo, Semantic Social Network Analysis, Ph.D. thesis of INRIA Sophia Antipolis – Mediterranee, Orange Labs, Telecom ParisTech. 2011.

[Faatz *et al*., 2002] A. Faatz, R. Steinmetz, Ontology enrichment with texts from the WWW, Acte de Conférence : *Proceedings of the 2nd Semantic Web Mining Workshop at ECMLI/PKDD*, Helsinki, Finland, 2002.

[Gandon, 2006] F. Gandon, Le Web sémantique n'est pas antisocial. 17èmes Journées Francophones d'Ingénierie des Connaissances, Nantes, France. 2006. Pages 131–140.

[Giunchiglia, 2008] F. Giunchiglia, Dynamic Ontology Matching : a survey. Technical Report DIT-06-046, Ingegneria e Scienza de l'Informazione, University of Trento, Italy. 2008.

[Groh et al., 2007] G. Groh, C. Ehmig, Filtrage Collaborative Vs. Filtrage Social, Group of Conference - Sanibel Island, Florida, USA, 2007

[Gruber, 2007] Thomas R. Gruber. Collective Knowledge Systems: Where the Social Web Meets the SemanticWeb. Journal of Web Semantics, Volume n°06, 2007, Leibniz.

[Hausenblas et al., 2008] M. Hausenblas, W. Halb et Y. Raimond, Scripting User Contributed Interlinking, 2008.

[Hernandez *et al.*, 2007] N. Hernandez, J. M. C. Chrisment, *Modeling context through domain ontologies, Journal of Information Retrieval, Contextual Information Retrieval Systems*, vol. 10, n° 2, avril 2007.

[Hurtado, 2008] E.V. Hurtado, ENVIRONNEMENTS COMMUNICANTS : « Une Réflexion sur La scénographie Événementielle », Etude de Master (Spécialisé Création en Nouveaux Médias), ENSCI, Ecole Nationale Supérieure de Création Industrielle, 2008. Paris.

[Jorio *et al.*, 2007] L. D. Jorio, L. Abrouk, C. Fiot, D. Hérin, M. Teisseire, Enrichissement d'ontologie basé sur les motifs séquentiels, *Actes de la Plateforme AFIA 2007, Atelier Ontologies et gestion de l'hétérogénéité sémantique*, Grenoble, France. 2007.

[Khoshgoftaar, 2009] M. Khoshgoftaar. Aperçu sur les techniques de filtrage collaboratif, Acte de Conférence : Conference of Advances in Artificial Intelligence, New York – USA –, 2009. Page 2 – 3.

[Kiryakov et al., 2004] A. Kiryakov, A. Kirilov, B. Popov, D. Manov, D. Ognyanoff, M. Goranov. Kim - semantic annotation platform. Review: *Journal of Natural Language Engineering*, Vol. 10, No. 3-4, 2004. Bulgaria.

[Lacombe et al., 2011] R. Lacombe, P. Bertin, F. Vauglin, A Vieillefosse, *Les données publiques au service de l'innovation et de la transparence*, Rapport remis le 13 juillet 2011 au Ministre de l'Industrie, de l'Énergie et de l'Économie numérique. Juillet 2011. Paris.

[Latanya Sweeney, 2002]. Latanya Sweeney, *Computer World:* Privacy algorithms, Octobre 2002.

[Laublet, 2008] P. Laublet, A. Passant. *Meaning Of A Tag : A collaborative approach to bridge the gap between tagging and Linked Data*. In Proceedings of the Workshop Linked Data on the Web, Karlsruhe, Germany, 2008.

[Leroy, 2010] V. Leroy, *Décentralisation des applications sociales*, Thèse de doctorat de l'INSA de Rennes**,** Institut de recherche en informatique et systèmes aléatoires de Rennes, 2010. France.

[Limpens, 2010] F. Limpens, Multi-points of view semantic enrichment of folksonomies, Thèse de Doctorat en Sciences, de l'Université de Nice - Sophia Antipolis (Mention Informatique), 2010. France.

[Mathes, 2004] A. Mathes, Folksonomies : Cooperative Classification and Communication Through Shared Metadata, University of Illinois Urbana-Champaign, December 2004.

[Massa et al., 2004] P. Massa, B. Bhattacharjee, Systèmes de recommandation à base de confiance. Acte de Conférence : *On the Move to Meaningful Internet Systems 2004: CoopIS, DOA, and ODBASE*, Berlin, Heidelberg, Springer, 2004. Pages 3-17

[Melville et al., 2002] P. Melville, R. Mooney. Content boosted collaborative filtering for improved recommendations, Acte de Conférence : *National Conference* on *Artificial Intelligence,* Acapulco, Mexico, 2002. Pages 189–192.

[Mika, 2007] P. Mika. Microsearch : An Interface for Semantic Search. Acte de Conférence : In Proceedings of the Workshop on Semantic Search (SemSearch 2007) at the 4th European Semantic Web Conference (ESWC 2007), volume 334 de CEUR Workshop Proceedings, in the Tyrol region of Innsbruck, Austria – Autriche – . Juin 2007.

[Neshatian *et al.*, 2004] K. Neshatian, M. R. Hejazi, *Text categorization and classification in terms of multi-attribute concepts for enriching existing ontologies*, Acte de conference : *Proceedings of the* 2*nd Workshop on Information Technology and its Disciplines*, Kish Island, Iran, 2004.

[O'Reilly, 2004] T. O'Reilly. What is web 2.0 ? design patterns and business models for the next generation of software. Acte de Conférence : Première conférence sur le Web 2.0, USA, 2004.

[Pasha, 2010] M. Pasha, *Grille sémantique autonome : Un intergiciel pour l'interopérabilité d'agents et services web*, Thèse de doctorat de l'université Européenne de Bretagne Sud, 2010. France.

[Passant, 2009a] Alexandre Passant, XSPARQL : Rapport technique, 2009, France.

[Passant et al., 2008] Alexandre Passant et Philippe Laublet (2008c). Ontologies et Web 2.0. Acte de Conférence : In IC2008, 19èmes Journées Francophones d'Ingénierie des Connaissances, au LORIA, Nancy, 2008.

[Passant, 2009] Alexandre Passant, Technologies du Web Sémantique pour l'Entreprise 2.0, Thèse de l'Université Paris IV, Sorbonne, 2009.

[Pazzani, 1999] M. Pazzani. A framework for collaborative, content-based and demographic filtering. Acte de Conférence : Artificial *Intelligence Review*, Volume 11, Number 1-5, 1999. Pages 13- 15.

[Raimond et al., 2008] Y. Raimond, C. Sutton et M. Sandler, Automatic Interlinking of Datasets on the Semantic Web, 2008.

[Rao et al., 2008] K. Nageswara Rao and V. G. Talwar. Application domain and functional classification of recommender systems a survey, Article de Revue: DESIDOC Journal of Library Information Technology, Section of Computer and Information Science. 2008. Pages 17–20.

[Rosnay, 2000] Joël de Rosnay, *L'homme symbiotique*. Nouv. éd. Paris, Seuil. 2000.

[Troncy, 2009] R. Troncy, Explorer des actualités multimédia dans le web de données, Acte de conférence : CWI Amsterdam, Science Park, 2009.

[Schmidt, 2010]. Loïc Schmidt, Passage à l'échelle des intergiciels RFID pour l'informatique diffuse, Thèse Doctorat de l'Université Lille 1 (Spécialité Informatique), 2010. France.

[Tandabany, 2009], Sattisvar Tandabany, Dynamic Composition of Functionalities of Networked Devices in the Semantic Web, Thèse Doctorat de l'Université Joseph Fourier, Grenoble 1 (Spécialité Informatique), 2009. France.

[Thiam, 2010] M. Thiam, Annotation Sémantique de Documents Semi-structurés pour la Recherche d'Information, Thèse de Doctorat des Universités de Paris-Sud et Gaston Berger (Spécialité informatique), 2010. France.

[Weiss, 2010] S. Weiss, *Edition collaborative massive sur réseaux Pair-à-Pair*, Thèse de Doctorat de l'université Henri Poincaré (spécialité informatique), 2010, Nancy. France

[Zacklad, 2007] M. Zacklad. Classification, thésaurus, ontologies, folksonomies : comparaisons du point de vue de la recherche ouverte d'information. Congrès : Lors du 35ème Congrès annuel de l'Association Canadienne des Sciences de l'Information, Canada, 2007.

[Zacklad *et al.*, 2005] M. Zacklad, H. Zaher, J. Cahier, *Recherche d'information dans le Web socio-sémantique, publication scientifique, Technique de l'ingénieur, 2005, Troyes.*

[Zargayouna & Nazarenko, 2010] H. Zargayouna, A. Nazarenko, Evaluation of Textual Knowledge Acquisition Tools, Acte de Conférence : The International Conference on Language Resources and Evaluation, Valletta (Malta), May 2010.

[Ziegler *et al.*, 2007] C.Ziegler, N. Golbeck, Investigating interactions of trust and interest similarity, Acte de Conférence : *ACM SIGIR Semantic and Information Retrieval Workshop*, Sheffield, UK, 2007.