



**HAL**  
open science

## Méta-données et annotations dans le Web sémantique

Serge Garlatti, Yannick Prie

► **To cite this version:**

Serge Garlatti, Yannick Prie. Méta-données et annotations dans le Web sémantique. Revue I3 - Information Interaction Intelligence, 2004, Web sémantique, Hors-série, pp.24. hal-02142939

**HAL Id: hal-02142939**

**<https://hal.science/hal-02142939>**

Submitted on 29 May 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Méta-données et annotations dans le Web sémantique

Yannick Prié\*, Serge Garlatti§

\* LIRIS FRE 2672 CNRS – Université Claude Bernard Lyon 1  
[yannick.prie@liris.cnrs.fr](mailto:yannick.prie@liris.cnrs.fr)  
§ Laboratoire2 Département IASC GET - ENST Bretagne  
[Serge.Garlatti@enst-bretagne.fr](mailto:Serge.Garlatti@enst-bretagne.fr)

## Résumé

*Le Web sémantique repose sur des langages et une infrastructure dont l'objectif est de se donner la possibilité d'enrichir le Web actuel à l'aide d'informations dites « sémantiques », utilisables par des machines, qui faciliteront la recherche et l'usage de ressources Web (pages Web, images, services, etc.). Il s'agit donc d'associer à ces dernières des informations structurées descriptives sous la forme de méta-données (ou annotations). Plusieurs dimensions sont alors à considérer : les types de ressources, plus ou moins fragmentées, concernées par les annotations / méta-données; l'automatisation plus ou moins marquée de la mise en place de ces dernières, la structuration plus ou moins forte de leur « schéma », les tâches qu'elles soutiennent (ou sous-tendent), ou encore l'utilisation plus ou moins « intelligente » qui en est faite par les agents logiciels. Après une rapide discussion de la notion d'annotation / méta-données, nous en proposons un scénario d'utilisation conséquent, dans le domaine de l'enseignement à distance. Nous montrons ensuite que si la problématique en soi n'est pas nouvelle, et est déjà traitée dans de nombreux domaines, l'échelle du Web et le niveau d'intégration technologique nécessaire sont quant à eux novateurs. Nous passons ensuite en revue un certain nombre de systèmes liés aux annotations / méta-données pour le Web sémantique, et concluons sur les directions probables que va prendre la recherche, en insistant sur la nécessité d'intégrer des chercheurs de SHS dans la réflexion.*

**Mots-clés** : annotation, méta-données, web sémantique.

# **1. PRESENTATION ET IMPORTANCE DE LA PROBLEMATIQUE DU POINT DE VUE DES USAGES**

## **1.1. Méta-données et annotations sur le Web sémantique**

Un des grands principes du Web sémantique est qu'il est nécessaire d'associer aux ressources du Web des informations exploitables par des agents logiciels afin de favoriser l'exploitation de ces ressources [6,8].

Associer par exemple une notice comprenant des champs : Auteur, Date de création, Date de modification, Mots-clés, à une page Web permet de considérer celle-ci non plus seulement comme comprenant du texte qui ne pourra qu'être traité statistiquement par un robot indexeur, mais également des informations structurées à la sémantique connue et utilisable comme telle par un agent logiciel. De la même manière, si ce qui apparaît comme un simple nombre dans une page Web est de façon explicite marqué comme un couple (valeur, devise), alors un agent pourra faire usage de ces connaissances.

Associer une information exploitable à une ressource signifie deux choses essentielles.

La première est que cette information doit d'une manière ou d'une autre être structurée – utilisable – et descriptive – de la ressource, de son utilisation. Il s'agit d'en faciliter et améliorer l'accès dans le cas d'une ressource directement visualisée par un utilisateur (par exemple en permettant une recherche d'information plus efficace et plus ciblée), mais aussi l'exploitation dans le cas d'une ressource exploitée dans le cadre d'un service à l'utilisateur (l'utilisateur n'est alors pas forcément conscient de l'utilisation de la ressource).

La seconde est que la ressource en question doit exister et pouvoir être exploitée sur le Web indépendamment des informations qui lui sont associées dans le cadre du Web sémantique : celles-ci sont utiles, mais non nécessaires pour accéder et utiliser la ressource, la page Web ou le service.

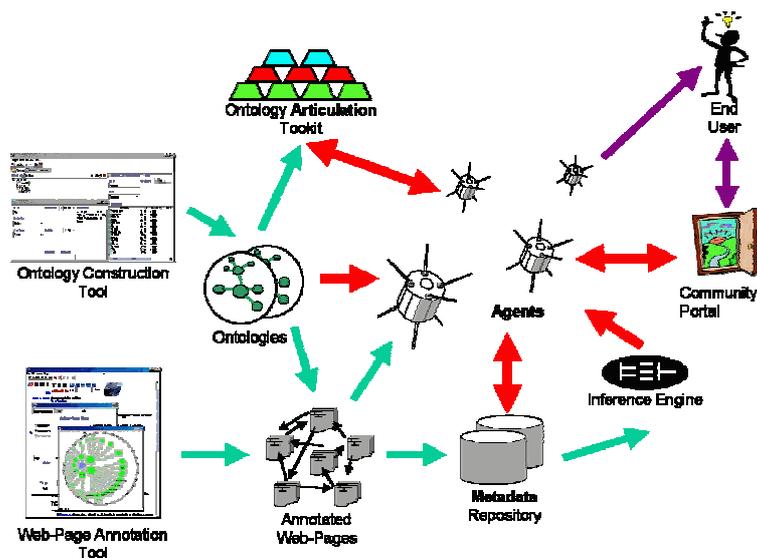


FIG. 1 – Une vision du Web sémantique

La figure 1 donne une vision (proposée sur le site <http://www.semanticweb.org/>) de l'utilisation des méta-données sur le Web sémantique. Des pages Web sont annotées à partir de connaissances disponibles dans une ou plusieurs ontologies (qui ont pour objectif de normaliser la sémantique des annotations), et ces annotations, regroupées en entrepôts de méta-données deviennent utiles pour des agents de recherche d'information, faisant ou non appel à des moteurs d'inférence permettant de déduire de nouvelles connaissances formelles des annotations.

Au-delà de ce schéma (sur lequel nous reviendrons au travers d'un scénario d'utilisation dans le domaine de l'apprentissage à distance), remarquons que deux termes principaux sont utilisés dans la littérature afin de décrire les informations associées à des ressources : méta-données et annotations.

Si nous revenons sur ces notions, et de façon générale :

- une méta-donnée est « une donnée sur une donnée ». Cette définition est un peu vague voire ambiguë, et elle est comprise de manière différente par différentes communautés qui conçoivent, créent,

décrivent, préservent et utilisent des systèmes d'information et des ressources. Par exemple, dans certains cas la donnée sur laquelle la méta-donnée porte est considérée comme ayant le même statut de donnée formalisée, traitable par un système informatique, dans d'autres, la donnée n'est qu'interprétable par un être humain, et seule la méta-donnée en permet le traitement automatique.

- une annotation est à la base une note critique ou explicative accompagnant un texte, et par extension, une quelconque marque de lecture portée sur un document, que celui-ci soit textuel ou image.

On le voit, les termes de méta-donnée ou d'annotation prennent bien en compte cette notion d'ajout d'information à une ressource, et on pourra a priori les utiliser indifféremment pour décrire ces informations que le Web sémantique doit ajouter au Web pour le rendre plus utilisable par des machines.

Pour autant, si ces deux termes existent, c'est qu'ils n'ont pas le même sens.

La communauté anglophone du Web sémantique, circonspecte sur la différenciation, considère par exemple dans le schéma que nous reprenons plus haut que les annotations de pages Web deviennent des méta-données dès qu'elles sont stockées dans une base sur un serveur. On différencie alors l'information en tant qu'elle est attachée à (et présentée avec) une ressource et l'information manipulable et exploitable de façon plus indépendante de celle-ci.

D'un point de vue plus lié à la pratique de l'annotation / méta-données (du point de vue du processus de mise en place), et s'il est nécessaire de distinguer, nous proposons ce distinguo :

- une méta-donnée sera plutôt attachée à une ressource identifiée en tant que telle sur le Web – aura plutôt une pertinence a priori et sera plutôt saisie suivant un schéma. Par exemple, la description normalisée d'un service Web, l'auteur d'un document, qui permettront de mettre en place des inférences.
- une annotation sera plus située au sein de cette ressource et écrite au cours d'un processus d'annotation / lecture. Par exemple, un commentaire libre associé à un fragment d'une page Web – quelques mots, un paragraphe – déterminé au besoin.

Evidemment, cette distinction n'a rien de définitive, il s'agit simplement de mettre l'accent sur le caractère plus situé au sein de la ressource (du fait de son exploitation par un utilisateur) de l'annotation, par rapport à une méta-donnée plus indépendante, voire ressource en tant que telle.

Plusieurs critères, non indépendants – que nous illustrerons dans la suite – peuvent être envisagés pour considérer les différents types de méta-données (annotations) :

- les types de ressources qu’elles concernent, plus ou moins fragmentées,
- l’automatisation plus ou moins marquée de leur mise en place,
- la structuration plus ou moins forte de leur « schéma »,
- les tâches qu’elles soutiennent (ou sous-tendent),
- l’utilisation qui en est faite par les agents logiciels, plus ou moins « intelligente ».

Remarquons que pour le Web sémantique, on parlera le plus souvent de méta-données ou d’annotations sémantiques. Deux possibilités d’interprétation sont ici disponibles : « sémantique » au sens de « sémantique formelle », et puis au sens plus vague de « faisant sens dans le cadre d’une tâche pour un utilisateur ». Si certains tenants du Web sémantique souhaitent limiter la portée de celui-ci au premier sens – formel – ces deux interprétations nous semblent pourtant tout à la fois pertinentes et de toutes façons inextricablement reliées dans le cadre d’un Web sémantique riche et ouvert, à l’image de la richesse et de la diversité du Web actuel, qu’on pourrait qualifier d’ « un peu sémantique ».

## **1.2. Exemples d’utilisation de méta-données et d’annotation et d’adaptation dans le Web sémantique**

### **1.2.1. Introduction**

Nous présentons dans la suite un scénario d’utilisation du web sémantique articulé autour de l’enseignement à distance. Ce scénario s’intéresse autant à l’utilisation des méta-données qu’à celle de l’adaptation qui est présentée dans un autre article de ce numéro.

« Le professeur Bern de l’université de Stuttgart est impliqué dans la création de cours diplômant sur les réseaux à l’intention d’étudiants de toute la Communauté Européenne. Afin de constituer son cours, il va chercher à utiliser toutes sortes de ressources trouvées sur le Web :

- Des ressources « classiques » telles que des articles scientifiques ou des cours numérisés : cours donnés en présence d’élèves et mis à disposition sous différents formats (textes, vidéos, présentations multimédia).

- Des ressources composées d'éléments proposés par différentes universités ou écoles, structurées par un ou plusieurs professeurs pour être accessibles en ligne, qui constituent des parcours cohérents à destination des étudiants. Certaines de ces ressources sont prévues pour être composées automatiquement par un logiciel à partir d'une spécification afin d'engendrer dynamiquement un cours personnalisé.

Pour cela, le scénario se divise en trois phases : la première est essentiellement centrée sur la recherche d'information principalement interprétable par un humain, la constitution d'un corpus et son indexation pour créer des cours en ligne manuellement. La seconde phase fait appel à un système de composition dynamique de cours adaptatifs spécifiés par un auteur ainsi qu'à l'utilisation de méta-données sémantiques interprétables par un logiciel. La troisième phase met en scène les apprenants.

### **1.2.2. Première partie : recherche d'information et création manuelle de cours**

Dans un premier temps, le Pr. Bern utilise un moteur de recherche de type « Google » pour retrouver des ressources intéressantes pour son cours. Pour cela, il commence par une requête simple : « Network » AND « Course », et obtient quelques millions de réponses, dont un grand nombre sont – dès les premières pages – non pertinentes. En effet, le terme « Network » y prend une sémantique bien différente que celle cherchée (réseaux sociaux, réseaux d'excellence, réseaux d'entreprise) tandis que « course » est présent dans de nombreuses pages dans l'expression « of course ». Découragé par la masse de document à analyser, le Pr. Bern décide de changer de méthode et d'utiliser un moteur de recherche fondé sur une approche de type « Web sémantique » utilisant un schéma de méta-données (Dublin Core – DC). L'outil lui propose un ensemble de descripteurs dont il peut préciser les valeurs ou domaines de valeurs à l'aide de mots clés. Par rapport à la recherche précédente, il peut désormais préciser si les termes recherchés sont dans le *Titre* du document (network), son *Sujet* (network) ainsi que sa *Date de création*. Par contre, le *Format* (html, pdf, etc.), le *Type* (course) et le *Langage* utilisé (anglais ou allemand) sont des contraintes disponibles dans son profil utilisateur. Il n'a donc pas à les fournir au système. En spécifiant ainsi un rôle et une sémantique pour les différents termes de sa requête, il lui est possible de mieux cibler sa recherche, et d'obtenir des ressources plus pertinentes. Cela n'est évidemment possible que parce que le moteur, lorsqu'il parcourt le Web utilise les méta-données Dublin Core associées aux ressources. Cependant, les utilisateurs d'un schéma de

méta-données tel que le Dublin Core, s'ils sont à peu près tous d'accord sur ce que signifie *Titre*, ne le sont pas forcément en ce qui concerne le *Sujet* ou le *Type* d'un document. Il pourra donc y avoir des ambiguïtés sur le rôle et le sens de ces champs.

Le Pr. Bern a quand même réussi à récupérer bon nombre d'éléments potentiels de cours qu'il va lui être possible d'organiser pour produire un cours complet. Comme à chacune de ses navigations sur le Web, il a ajouté des annotations textuelles sur les parties les plus pertinentes de chaque ressource (ensemble de paragraphes, images), décrivant la ou les utilisations qu'il prévoit d'en faire plus tard, ce qui facilitera sa tâche ultérieure.

Conformément à la pratique de son institution, le Pr. Bern ajoute tout d'abord les documents pertinents à la base générale documentaire de l'Université. Pour cela, il va être obligé de modifier certains champs de méta-données. A la différence de la pratique habituelle sur le Web, un thésaurus propre à l'Université est utilisé. Celui-ci décrit le vocabulaire à utiliser dans les différents champs, et sert de guide d'utilisation du schéma. De par le référentiel commun constitué par le thésaurus, le Pr. Bern et ses collègues sont à peu près sûrs d'utiliser les mêmes termes pour décrire les documents, et peuvent utiliser la structure de navigation constituée par le thésaurus pour atteindre ceux-ci. Pour toutes les saisies de méta-données du DC, le Pr. Bern n'est pas obligé de remplir tous les champs puisque certains sont obtenus à partir de son identifiant personnel et des références de son université tels que « créateur, éditeur, source, date, format ». Il est par contre bien évident qu'il doit remplir lui-même les champs « titre, sujet, couverture, relation, droits, langage et description ».

Un des documents ajoutés à la base possède dans sa description Dublin Core un élément qui va être utilisé automatiquement par un logiciel d'alerte, permettant de prévenir trois abonnés de l'Université qu'ils peuvent consulter un article les intéressant potentiellement. L'un d'eux va le consulter et y laisser quelques annotations de lecture, car il n'est pas du tout d'accord avec une affirmation de la troisième partie. Une discussion s'engage alors lorsque les deux autres utilisateurs ainsi que le Pr Bern (prévenu que le document a été annoté) s'opposent à cet avis avec véhémence. Chacun faisant référence dans son argumentation à d'autres parties du document.

Afin de mettre en place le cours lui-même, le Pr. Bern va réutiliser les documents. Certains sont réutilisables en l'état (par exemple une figure illustrant particulièrement bien une notion), et d'autres ne le sont pas. En

effet, sortis de leur contexte de création, ils ne sont pas intégrables directement dans un parcours cohérent, et demandent une fragmentation / modification / adaptation, ainsi qu'une organisation *ad hoc*.

Si la précision d'utilisation avec le moteur de recherche utilisant les méta-données est meilleure qu'avec le moteur classique, et s'appuie sur des méta-données mises en place de façon simple et peu coûteuse en temps, elle n'est suffisante que pour une utilisation par un utilisateur humain, et non par un logiciel de génération automatique de cours. En effet, il est uniquement possible de l'utiliser pour de la recherche d'information qui sera interprétée (et exploitée) par un humain : même si un schéma tel que le Dublin Core définit une structure de méta-données, l'exploitation de celle-ci par des logiciels est limitée puisque aucune sémantique – interprétable par une machine – n'est associée aux domaines de valeur des différents champs et donc certaines ressources retrouvées pourraient ne pas convenir. Même si les termes sont les bons dans le bon rôle, ils peuvent ne pas avoir été utilisés avec une même sémantique, ou encore la qualité de la saisie des méta-données n'a pas été validée ou vérifiée.

### **1.2.3. Seconde partie : composition dynamique de cours adaptatifs**

Pour composer son cours, le professeur Bern utilise un nouvel environnement appelé « E-Learning 2010 ». Ce dernier engendre dynamiquement un cours adapté à chaque étudiant (connu d'après son profil d'étude). « E-Learning 2010 » calcule un cours adapté à chaque étudiant à partir d'une structure de cours choisie par un professeur et d'une spécification du contenu selon un schéma de méta-données appelé « LOM+ », ainsi que des modes d'évaluation des contenus et de leur utilisation par les méthodes d'adaptation. LOM+ utilise des ontologies comme vocabulaire partagé pour les termes du domaine du cours (ici les « réseaux informatiques ») ; pour les modèles de cours en fonction d'une approche pédagogique choisie (une structure globale de cours qui organise l'accès aux contenus qui sont typés : présentation d'une notion, puis exercice, approfondissement, etc.) ; enfin pour les modes d'évaluation et les méthodes d'adaptation pour un modèle utilisateur donné.

Toutes les ressources qu'il a trouvées / définies précédemment peuvent être réutilisées, mais il est nécessaire d'y associer de nouvelles méta-données LOM+ pour leur permettre d'être insérées dans la trame du cours. La saisie des méta-données à partir de « E-Learning 2010 » est

bien plus simple que précédemment puisque des domaines de valeurs – définies par différentes ontologies (domaine, approche pédagogique, modèle utilisateur, adaptation) – sont associées aux différents champs. Pour chaque champ, l'outil propose soit une liste de termes, soit une valeur déduite par le système en fonction du contexte, qu'il suffit de valider ou non selon le type de champ (pour le champ « Status », dans la catégorie « Lyfecycle » la valeur est directement proposée par le système à partir du contexte courant, par contre pour les champs « Difficulty » et « IntendedEnduserRole », les valeurs suivantes sont respectivement proposées : <low, average, high> and <PassiveLearner, ActiveLearner, CollaborativeLearner, TutorLearner, ProfesseurLearner>). Sont fournis également un guide d'utilisation du champ munis d'exemples et de contre-exemples et un navigateur d'ontologie permettant de mieux comprendre la structure de celle-ci et donc la sémantique des concepts ou relations liés à un terme donné. On voit bien dans l'exemple précédent que sans explication ou guide d'utilisation, il aurait été difficile au professeur Bern de fixer une valeur pour ces champs.

Il est possible de rechercher plus précisément de nouvelles ressources qui sont disponibles à partir de « E-Learning 2010 ». Pour cela, il suffit au Professeur Bern de sélectionner un élément du modèle de cours muni d'un type – exercice, solution, problème, définition, approfondissement, explication, éclaircissement, etc. – et une interface de recherche d'information est proposée n'utilisant que les champs de méta-données pertinents dont certains sont déjà définis comme par exemple le type d'élément. Pour les autres une liste de valeurs est disponible. S'il n'est pas possible au système de trouver dans la base de « E-Learning 2010 », les ressources nécessaires, il lui est possible de réaliser une recherche élargie sur le Web. Cette recherche étant guidée par les ontologies disponibles, il n'y aura plus d'ambiguïté sur les sens des termes utilisés. En effet, il est possible au logiciel de recherche de rechercher toutes les ressources indexées à partir de ces mêmes ontologies ou d'une partie d'entre elles ou encore d'élargir à d'autres ontologies équivalentes mais avec lesquelles il sera nécessaire de réécrire les termes pour assurer une transcription jugée sémantiquement juste. Il lui est également possible de fixer certains critères d'évaluation dans son profil utilisateur pour une recherche adaptative afin de filtrer les résultats ou de le guider dans sa recherche.

Après avoir choisi une approche pédagogique, le professeur Bern sélectionne une structure de cours particulière qui convient à sa manière d'enseigner. Il obtient ainsi une structure globale de cours dans laquelle il doit spécifier des contenus à partir du schéma de méta-données « LOM+ ». Dans ce schéma de méta-données, certaines entrées sont

dédiées à l'adaptation tels par exemple les pré-requis qui prennent leurs valeurs dans l'ontologie du domaine, le type d'étudiants – 1<sup>re</sup> année, 2<sup>e</sup> année, etc. C'est cette spécification qui permettra alors au système de générer les cours en fonction des profils des étudiants - qui ont parmi leurs caractéristiques utilisateur un modèle de recouvrement permettant de connaître pour chaque concept du modèle de domaine le niveau de connaissance acquis ou supposé acquis par l'étudiant. Pour l'adaptation, le Pr. Bern doit aussi préciser pour différents stéréotypes d'étudiants les méthodes d'adaptation autorisées, les modes d'évaluation des contenus. Chaque méthode d'adaptation utilise le résultat de l'évaluation des fragments.

Pour la constitution du cours, le professeur Bern a parfois eu des difficultés à bien comprendre la méthodologie de conception liée à « E-Learning 2010 », car cet environnement est assez complexe. Cependant, en plus de la documentation des diverses ontologies et schémas de méta-données, une assistance logicielle et humaine tant au niveau interne qu'externe lui est proposée pendant la conception. Sa tâche est à la fois rendue plus complexe sur certains aspects mais aussi grandement facilitée pour d'autres. Ses cours ou éléments de cours sont plus facilement réutilisables pour lui-même et pour les autres puisqu'il lui a fallu avoir une réflexion sur la granularité – taille de chaque élément et niveau d'abstraction pour ceux qui ne sont pas atomiques – des éléments et leur réutilisation. Un suivi de la qualité des méta-données est mis également en place au sein de l'Université afin d'assurer la cohérence de chaque cours ou élément de cours disponible en ligne. Ces dernières doivent être saisies de manière bien plus rigoureuse lorsqu'une partie des champs est destinée à un logiciel de composition de cours. En effet, si l'intention de l'auteur lors du choix d'une valeur de champ n'est pas conforme à l'utilisation que va en faire le logiciel de composition, le résultat ne sera guère prévisible

#### **1.2.4. Troisième partie : lecture et apprentissage par un apprenant**

Chaque étudiant accédant à « E-Learning 2010 » peut choisir un cours et une approche pédagogique particulière (par résolution de problème, collaborative, etc.). En fonction de son niveau scolaire et de ses acquis précédents, certains cours complémentaires – unités de valeurs – peuvent lui être proposés à partir de son profil utilisateur. En effet, ces unités de valeurs peuvent être indispensables à une bonne compréhension. Dans certains cas, ces différents éléments peuvent aussi résulter d'une négociation avec son tuteur afin par exemple d'assurer une charge de

travail raisonnable, ainsi qu'une progression satisfaisante de son apprentissage. Notre étudiant étant à l'université du Sussex, il aura à sa disposition une présentation et une structure de ses pages de cours – contenu, barre de navigation et fonctions - qui sont communes à tous les étudiants du Sussex. En effet, ceux de l'université de Stuttgart n'ont pas tout à fait les mêmes habitudes, et ont donc un environnement légèrement différent. Chaque étudiant peut potentiellement disposer d'une structure globale de cours et d'un contenu différent. Néanmoins, chacun peut accéder à l'ensemble s'il le désire.

Chaque étudiant se doit d'être au courant de son propre modèle utilisateur et de ses parties privées / publiques. Un étudiant possède un modèle utilisateur global qui peut être enrichi dynamiquement en fonction des cours auxquels il est inscrit. Pour chaque cours, avant d'accepter l'enrichissement de son modèle, une négociation a été faite entre le système, le tuteur et l'apprenant afin de se mettre d'accord sur les données privées / publiques. En effet, certaines sont obligatoires pour le suivi de l'apprenant par un tuteur et d'autres peuvent être gardées privées ou publiques – pour partager avec d'autres apprenants par exemple. Tout au long de l'apprentissage, ce modèle global est maintenu et enrichi au fur et à mesure de la progression de l'étudiant dans ses études. Comme ces informations sont disponibles sur un serveur de modèles utilisateur, il lui est tout à fait possible de continuer ses études dans un autre pays et de continuer à utiliser son modèle. Ceci n'est possible que parce que des vocabulaires communs – ontologies – sont partagés entre différents pays.

On le voit dans cet exemple, si le simple ajout de méta-données à une ressource (document ou fragment) n'est le plus souvent qu'orienté vers une tâche de « recherche d'information » dans toute sa généralité (schéma général), des schémas de description des ressources peuvent être spécialisés pour des tâches (lecture active) ou des domaines d'application particuliers. Par exemple la construction ou l'utilisation de cours en ligne, les échanges liés à l'information (XMLNews, flux RSS) ou aux produits bancaires, les documents audiovisuels (MPEG7). Une même ressource peut bien entendu être décrite suivant plusieurs espaces d'indexation (points de vues). Ces schémas de méta-données se résument le plus souvent en des standards descriptifs exprimés sous la forme de DTD ou de schéma XML qui dans de nombreux cas, n'utilisent bien souvent que des mots-clé, voire au mieux des vocabulaires standardisés tels que AAT (Art and Architecture Thesaurus), TGN (Thesaurus of Geographic Names)<sup>1</sup>. Ils ne permettent alors ni de définir de manière précise la

---

<sup>1</sup> Voir [http://www.getty.edu/research/conducting\\_research/vocabularies/download.html](http://www.getty.edu/research/conducting_research/vocabularies/download.html) pour ces deux vocabulaires.

sémantique opérationnelle (liée à des inférences automatiques) des différents champs des descriptions, ni celles de leurs domaines de valeurs. L'étape suivante proposée par le Web sémantique est donc celle des schémas de méta-données fondés sur des ontologies, autorisant des inférences formelles sur les méta-données ou les annotations.

Le scénario que nous venons de présenter illustre un certain nombre de points importants liés aux annotations / méta-données :

- Il est possible d'annoter une ressource dans son ensemble, ou un fragment de ressource documentaire (pour peu que la ressource soit fragmentable : une image ou un texte l'est, un service de réservation d'hôtel l'est moins immédiatement).
- Les schémas de méta-données / annotation peuvent être plus ou moins généraux ou spécialisés : du Dublin Core permettant de décrire tout type de documents avec des descripteurs basiques, à une ontologie formelle permettant d'annoter précisément des fragments de cours, ou des dépêches d'agence pour la veille financière (rachat, annonce de perte, etc.), beaucoup de choses sont possibles.
- La mise en place des annotations peut être plus ou moins automatisée : de la mise en place manuelle (« tel paragraphe est intéressant, je vais le réutiliser tel quel »), à l'automatique (cette suite de chiffres suivie du caractère € signifie qu'il convient de l'annoter avec le concept *Prix*), en passant par le semi-automatique (je vous propose de considérer que ce « 37000 » est un code postal, à vous de me le confirmer).
- Les méta-données / annotations peuvent être utilisées ou connues des utilisateurs finaux (cas des données permettant de rechercher avec le Dublin Core), ou bien uniquement par des spécialistes (les concepts de description d'une progression de cours ne sont connus que des enseignants).
- Les méta-données / annotations sont utilisées pour deux grandes tâches. La première est la recherche d'information, puisque toute méta-donnée informatique liée à une ressource représente de fait un index pour cette ressource, et peut être utilisée comme telle. La seconde est la composition de documents, de fragments de documents, de services en vue de construire de nouveaux documents (cours à la carte, catalogues, réponses à une question, etc.) ou de nouveaux services (hôtel + spectacle).
- Les schémas de méta-données sont plus ou moins formalisés et permettent d'offrir des services considérés comme plus ou moins intelligents, puisque les agents logiciels peuvent plus ou moins s'en emparer pour mener des inférences.
- Les annotations et méta-données peuvent rester dans la sphère personnelle (mes annotations sur une page Web), ou s'inscrire dans

une collectivité plus ou moins restreinte (mon groupe de travail, mes étudiants, l'ensemble des usagers du Web). L'annotation peut être mise en place de façon collaborative, être plus ou moins partagée ou publiée. On peut avoir plus ou moins confiance en des méta-données.

- Point plus technique : les méta-données / annotations de ressources du Web peuvent être stockées ou mises à disposition dans ou hors de cette ressource : les documents peuvent être « équipés » par leurs méta-données (alors un moteur pourra collecter celles-ci), ou bien on pourra avoir de véritables bases de (méta-) données centralisant celles-ci. La diffusion des méta-données peut se faire de façon centralisée, de façon plus directe (P2P), en faisant appel à des proxys, etc.

## **2. METHODES, TECHNIQUES, OUTILS EXISTANTS SUR LESQUELS ON PEUT S'APPUYER**

Il va de soi que le Web sémantique, considéré du point de vue des annotations et des méta-données ne peut que s'appuyer sur les standards existants qui forment l'ossature technologique du Web :

- protocole HTTP pour la transmission,
- langages HTML et XML, feuilles de style pour la présentation des résultats,
- langage RDF (Resource Description Framework) pour l'expression et l'échange des méta-données elles-mêmes, sous la forme de triplets, mêmes si ceux-ci peuvent également être stockés dans des bases de données pour une gestion plus efficace.

De la même façon que le Web n'a pas attendu le Web sémantique pour offrir des informations qualifiables de « sémantiques », les problématiques des méta-données et des annotations sont prises en compte depuis que le Web est Web. Les méta-données sur des ressources Web sont utilisées dans tous les systèmes d'indexation de documents Web, que ce soient des moteurs de recherche, des annuaires, des signets personnels, et de façon plus générale dans tous les systèmes de gestion d'information, ces derniers bénéficiant d'une vieille tradition dans le monde de la documentation et des bibliothèques (schéma d'indexation, utilisation de thésaurus, etc.). De la même manière, les échanges électroniques de documents au sein de communautés plus ou moins fermées à travers le réseau ont souvent conduit celles-ci à définir et à utiliser de façon courante aussi bien des modèles de documents que des schémas de méta-données adaptés à leurs pratiques, conduisant à de véritables normes. L'utilisation de thésaurus et autre modèles de

connaissances afin d'indexer des documents au sein d'intranets, dans le cadre de la « gestion des connaissances » est également une réalité sur laquelle le Web sémantique ne peut faire l'impasse.

Depuis le fameux navigateur Mosaic, déjà équipé en 1994 de la possibilité de gérer des annotations, de nombreux systèmes ont vu le jour, qui permettaient à des utilisateurs d'annoter des documents du Web, et de lire des annotations. En ce qui concerne le stockage des annotations, beaucoup de possibilités ont été essayées : dans les documents eux-mêmes (balises ad-hoc, RDF), dans les URL, dans des bases présentes sur des serveurs centralisés. Ces annotations, le plus souvent sous forme simplement textuelle, éventuellement typées (commentaire positif, négatif, voire lien – on peut alors parler d'« annotation-lien » –, etc.) ont pu être plus ou moins partagées. De nombreux systèmes – Yawas [3], Critlink [11], etc. –, y compris commerciaux (comme le défunt Thirdvoice) ont été lancés, souvent avec des résultats décevants : pas de diffusion généralisée des principes. L'expérience accumulée au cours de ces multiples essais mérite sans aucun doute d'être analysée dans le cadre du Web sémantique.

On peut donc dire que ce sont tout à la fois :

- des méthodes de conception de schémas de méta-données (thésaurus, ontologies),
- des outils et des principes d'utilisation de méta-données, d'annotation, de présentation de celles-ci, aussi bien dans le monde numérique qu'à l'extérieur (par exemple [7] pour une revue des types d'annotation),
- des architectures de stockage / requêtes / diffusion d'annotations (voir [4] pour une discussion des différentes architectures),
- des modèles et des langages de descriptions et d'échange de méta-données / annotations,

qui constituent un référentiel aussi bien technique que méthodologique sur lequel s'appuyer.

Ajoutons à cette liste

- les multiples outils, méthodes et techniques d'extraction d'information de documents sur le Web (fouille de données textuelles, méthodes statistiques ou plus ou moins intelligentes, fondées sur des connaissances et spécialisées) propres à proposer à l'utilisateur telle annotation, ou à remplir automatiquement tel ou tel champ de méta-données.
- les outils et méthodologies permettant l'extraction et la mise en place plus ou moins automatisée de schémas de description à partir de

corpus Web (terminologies, ontologies formelles), la construction coopérative de ces schémas, etc.

- les outils d'échanges de contenu selon les principes du « peer to peer », et l'expérience massive d'ajout distribué de méta-données distribuées qui en résulte.

On le voit donc, l'approche de Web sémantique, si elle n'est pas à proprement parler nouvelle du point de vue des méta-données ou des annotations, y compris à base d'ontologies formelles, peut avec raison s'appuyer sur des nombreux travaux, techniques, modèles et outils. Cela est évidemment vrai du côté des techniques du Web (puisque c'est le socle du Web sémantique), mais aussi du côté de tous les domaines de recherche liés à la recherche et à la gestion d'informations et des connaissances.

L'approche du Web sémantique, qu'on pourrait qualifier de plus « systématique » pourra par contre contribuer à l'unification de nombre de ces travaux sous une bannière commune, du fait de la stabilisation et de la standardisation de langages et de principes architecturaux relativement similaires.

### **3. TRAVAUX ET RESULTATS EXISTANTS DU WEB SEMANTIQUE**

Il nous apparaît que les travaux sur le Web sémantique, considérés du point de vue des annotations et des méta-données peuvent se répartir comme suit :

- travaux sur les langages de description d'ontologies et de méta-données ;
- travaux sur la construction des ontologies pour associer des méta-données, décrire des pages Web ;
- travaux sur l'utilisation de méta-données pour la personnalisation de pages Web ;
- travaux sur des applications pilotes, mettant en jeu :
  - des schémas / ontologies pour annoter,
  - des outils pour annoter conformément à ces schémas,
  - des applications faisant usage des méta-données en tant que telles (recherche d'information), et/ou les utilisant pour les présenter aux utilisateur en même temps que les ressources annotées.
  - des architectures pour soutenir ces applications,
  - des « leçons apprises ».

Nous nous concentrerons dans cette partie uniquement sur ces derniers travaux (liés aux applications permettant de mettre en œuvre de façon concrète les grands principes du Web sémantique), les autres types de travaux étant traités dans d'autres parties du présent numéro.

Pour cela, nous présentons dans un tableau récapitulatif un état de l'art de quelques systèmes liés au Web sémantique, considérés comme représentatifs du fourmillement actuel<sup>2</sup>. Ces systèmes sont étudiés suivant différents critères :

- *Types de ressources annotées* : que peut-on annoter à l'aide du système ? (une page Web dans son ensemble, un fragment de document XML délimité par deux balises ?) ;
- *Langage pour les annotations / localisation* : dans quel langage de description sont représentées les annotations ? (texte simple, langage structuré comme XML, langage de description de triplets comme RDF ?). Quel est plus spécifiquement le moyen de désigner les fragments ?
- *Schémas de méta-données* : dans quel langage sont exprimées les contraintes sur les annotations que l'on peut créer ? Un ensemble d'attributs / types de valeurs (ex. Creator / chaîne de caractères) comme le Dublin Core ; un langage de description d'ontologies comme DAML+OIL ou OWL ?
- Architectures des systèmes : considérations techniques sur le stockage, la recherche et la diffusion des méta-données ;
- Utilisation actuelle possible des systèmes et des méta-données présentées.

---

<sup>2</sup> Ce tableau a été construit en avril 2003, légèrement actualisé en octobre 2004.

Système / application	Types de ressources annotées	Langage pour les annotations/localisation	Schéma de méta-données	Architecture	Utilisation des métadonnées	Remarques
Annotea <a href="http://www.w3.org/2001/Annotea/">http://www.w3.org/2001/Annotea/</a>	Fragments de pages Web HTML et XML (SVG, XHTML,...)	RDF / XPointer	Schéma RDF simple : une classe annotation, des propriétés titre, corps, auteur, date, etc.	Serveurs HTTP dédiés. Interface d'annotation. Navigateur instrumentés pour présenter les annotations	Affichage des annotations associées lors de la lecture suivante de cette page	Possibilité de répondre à une annotation par une autre. Diverses implémentations (mozilla, IE)
Photostuff <a href="http://www.minds.wap.org/2003/PhotoStuff/">http://www.minds.wap.org/2003/PhotoStuff/</a>	Images + fragments d'images	RDF / OWL	Ontologie OWL	Utilisation locale ou serveur Mindswap	Recherche d'images.	
SHOE Knowledge Annotator <a href="http://www.cs.umd.edu/projects/plus/SHOE/KnowledgeAnnotator.html">http://www.cs.umd.edu/projects/plus/SHOE/KnowledgeAnnotator.html</a>	Page Web HTML	Extension de HTML (Simple HTML Ontology Extension) pour décrire des ontologie / des annotations dans la page Web sous la forme d'assertions liées à l'ontologie	Ontologie SHOE : classification, relation, inférences.	Interface graphique pour récupérer les assertions, et les stocker dans une base (Pauka). Requêtes à cette base.	Recherche d'information.	Interface graphique de construction de requêtes. Export des ontologies SHOE vers DAML. Ancêtre.
COHSE Annotator <a href="http://www.ecs.soton.ac.uk/~rmb/colts/annotator/">http://www.ecs.soton.ac.uk/~rmb/colts/annotator/</a>	Fragments de pages Web HTML	RDF	Ontologie DAML	Outil pour annotation. Serveurs d'annotation dédiés.	Génération d'« hypermédias ontologiques » (pages Web enrichie par des annotations à base d'ontologies). Visualisation des A	Lié au projet COHSE Conceptual Open Hypermedia Project <a href="http://cohse.semanticsWeb.org/">http://cohse.semanticsWeb.org/</a>
WEBKB <a href="http://www.webkb.org/">http://www.webkb.org/</a>	Toute ressource Web	RDF	Ontologie – graphes conceptuels. Ontologie générale construite à partir de Wordnet.	Serveur partagé	Recherche d'information	Interface très complexe

Système / application	Types de ressources annotées	Langage pour les annotations/localisation	Schéma de méta-données	Architecture	Utilisation des métadonnées	Remarques
Ontomat Annotizer <a href="http://annotation.semanticweb.org/tools/ontomat">http://annotation.semanticweb.org/tools/ontomat</a> [5]	Fragment de page Web	RDF	Ontologie OWL	Outil de création de pages Web et annotation. MD dans les documents, robot + serveur inférentiel.	Toute utilisation : visualisation, recherche inférentielle.	Voir aussi OntoAnnotate (Outil d'annotation de Ontoprise). V3 en cours de développement
SMORE; Semantic Markup, Ontology and RDF Editor V3 <a href="http://www.mindswap.org/~ajitka/Editor2.shtml">http://www.mindswap.org/~ajitka/Editor2.shtml</a>	Fragments de pages Web HTML, courriels Images SVG	RDF	Schéma RDF. Fondé sur des ontologies extérieures a priori ou construction du schéma après annotation.	Serveur dédié	Création / annotation de pages Web. Portails liés aux annotations pour RI, gestion.	Version pour annotation d'image et de courrier électroniques. V4 en dévpt. compatibilité OWL prévue
Résumés annotés <a href="http://annotation.semanticweb.org/iswc/documents.html">http://annotation.semanticweb.org/iswc/documents.html</a>	Abstracts HTML pour la conférence ICWS 2002	Triplets RDF dans les abstracts	Schéma RDF / DAML simple : auteur, page Web, email, adresse, institution, date de création.	Utilisation de Webscripter pour parser les descriptions	Génération de site Web de résumés.	
RDF Web / FOAF <a href="http://rdfweb.org/">http://rdfweb.org/</a>	Photographies de personnes, pages web personnelles	RDF	Schéma RDF + propriétés DAML décrivant une personne sur le Web	Serveurs dédiés.	Recherche de personnes, de liens entre personnes, etc.	Possibilité d'annotation de fragments d'images SVG
MnM <a href="http://kmi.open.ac.uk/projects/akl/MnM/">http://kmi.open.ac.uk/projects/akl/MnM/</a>	Fragments de Pages Web HTML	RDF / DAML+OIL, OCML	Ontologies DAML+OIL ou Schéma RDF	Browser Web, outil de navigation d'ontologie et de création d'instance dans une BC	Toute utilisation : visualisation, recherche inférentielle	Annotation automatique des pages grâce au plugin <i>Amilcare</i> (analyse de la langue)

Au vu du tableau, qui décrit quelques grands types d'outils d'annotation ou de systèmes à base d'annotation, quelques commentaires peuvent être faits sur le panorama qu'il dresse :

- RDF est largement utilisé pour exprimer les méta-données associées à des ressources Web et joue donc le rôle pour lequel il avait été prévu.
- En ce qui concerne les schémas de méta-données, un certain nombre de systèmes se limitent à de simples schémas RDF (des classes, pas forcément en hiérarchie, auxquelles sont associées des propriétés), sans aller jusqu'à utiliser des langages spécialement conçus pour décrire des ontologies (DAML+OIL, progressivement remplacé par OWL).
- La gestion des méta-données d'un point de vue « inférentiel » se limite alors à leur utilisation par des programmes dédiés : alors qu'un champ *Creator* tel que défini par le Dublin Core ne dit absolument rien sur ce qu'il est possible d'en faire, avec OWL LITE, il est possible de préciser que deux instances d'un concept Créateur : *Romain Gary* et *Emile Ajar* sont par exemples considérées comme identiques, dans le langage d'expression de méta-données lui-même, donc utilisable par tout système connaissant ce langage.
- Les outils d'annotation permettent soit d'annoter des ressources en tant que telles (dans les faits, décrites par des URL), soit des fragments, pour deux types de documents (pages HTML, et XML en général avec XML-fragment / XPointer : exemple SVG).
- Les outils d'annotations permettent soit des annotations *a posteriori* sur les ressources, soit – pour des pages Web – la conception de celles-ci et l'intégration concomitante de méta-données (Cf. Smore et Ontomat).
- Les méta-données doivent être stockées sur des serveurs dédiés (le plus souvent serveurs Web instrumentés avec des services). Elles peuvent l'être directement, ou bien après passage d'un robot (crawler) parcourant le Web et les extrayant des pages Web rencontrées<sup>3</sup>.
- On assiste à deux principes de développement différents. Dans un cas, le problème est attaqué de façon globale, et est pensé pour tout faire d'entrée de jeu (KAON suite + Ontomat), dans un autre (Annotea), on a la mise en place d'une infrastructure minimale (architecture, schéma RDF) d'annotation de fragments sur le Web, puis extension pour prendre en compte plus de choses : passage à l'annotation d'image, extension puis changement de schéma, etc.

---

<sup>3</sup> Voir par exemple le moteur *Swoogle* de l'Université du Maryland <http://swoogle.umbc.edu/index.php>

Au niveau architectural, les directions prises par les travaux sur les méta-données sont liées à la distribution et au stockage des méta-données – voir par exemple l'étude de la distribution des méta-données RDF suivant les principes P2P (peer to peer) : les travaux autour d'EDUTELLA (<http://edutella.jxta.org/reports/edutella-whitepaper.pdf>), ou le projet SWAP (Semantic Web and Peer to peer <http://km.aifb.uni-karlsruhe.de/projects/swap>). Au niveau de la mise en place semi-automatique d'annotation, on citera l'outil d'extraction d'information AMILCARE (<http://nlp.shef.ac.uk/amilcare/>), impliqué dans un nombre respectable de projets.

#### **4. RECHERCHES FUTURES POUR LE WEB SEMANTIQUE**

Il nous apparaît que les travaux liés aux annotations / méta-données visant à mettre en action les théories du web sémantique sont nombreux, et que force est de constater qu'ils commencent à fournir des résultats techniquement impressionnants en terme de prototypes et d'architectures, concrétisant énormément d'efforts au niveau international (par exemple KAON / Ontomat). Cependant, même si les bénéfices potentiels apportés par ces nouvelles possibilités d'applications sont indéniables, il ne nous semble pour l'instant pas exister d'applications réellement utilisées de façon routinière sur le Web.

Deux explications peuvent être données à cet état de fait : (1) il est encore trop tôt pour que les prototypes sortent des laboratoires pour aller directement sur le Web (peut-être un passage par une case intranet est-il nécessaire pour rôder à petite échelle un certain nombre de techniques). (2) il ne suffit pas de bonnes idées et de leur mise en place en termes techniques pour entraîner l'utilisation des systèmes (problème somme toute classique en informatique).

On peut considérer plusieurs cas pour le statut des méta-données vis-à-vis des utilisateurs et des concepteurs qui les manipulent :

- Dans le cas d'applications pour lesquelles les annotations / méta-données sont connues et éventuellement mises en place par les utilisateurs finaux des systèmes, il nous semble qu'il faut qu'un certain nombre de conditions soient remplies afin que les systèmes développés rencontrent du succès, à savoir
  - un bon schéma de méta-données lié à la pratique d'une communauté,

- associé à la diffusion dans cette communauté d'outils facilement utilisables,
- lesquels apportent immédiatement un gain aussi bien individuel que collectif incitant les personnes qui le doivent à ajouter les méta-données aux ressources communes.<sup>4</sup>
- Dans un cas intermédiaire, celui où existent des concepteurs « secondaires », comme par exemple les enseignants qui spécifient des méta-données dans le cadre de la conception de cours en ligne adaptatifs, alors que les étudiants ne font qu'utiliser les documents qu'on leur présente, on se retrouve dans le premier cas, avec une communauté plus réduite.
- Enfin, dans le cas d'applications pour lesquelles les annotations / méta-données ne sont connues et mises en place que par les concepteurs des systèmes, c'est-à-dire lorsque l'utilisateur final n'est jamais conscient des mécanismes en jeu, on se retrouve dans le cas plus classique de développement de services sur le Web, qui peuvent fonctionner ou non.

Dans un Web sémantique concerné par l'instrumentation de ressources principalement documentaires à l'aide d'ontologies (que ce soient de simples schémas de méta-données, ou bien des hiérarchies de types et de concepts plus raffinées) se pose d'entrée de jeu la problématique de l'utilisation et de la compréhension de ces ontologies par les utilisateurs, qui doivent en acquérir une expertise suffisante.

Cette expertise passe par la compréhension de l'utilisation qui va être faite des méta-données, et donc des inférences qu'elles vont permettre, à divers degrés. Par exemple, si j'affirme qu'une chaîne de caractères est bien le *Titre* de mon document, je suppose qu'un outil permettra de mener une recherche dans ce champ seul, ou bien pourra construire une table des matières à partir de lui. Si j'affirme que ce paragraphe présente une *Notion* du cours que je prépare, que cette notion précède telle autre dans le déroulement du cours, et que ce document apporte des *Eclaircissements* (sachant qu'un *Eclaircissement* est une sorte d'*Explication*), alors je dois savoir quelle est l'utilisation qui sera faite de cette indication : par exemple génération d'un document listant les notions, génération d'un document composé de paragraphes se suivant dans un ordre adapté à l'étudiant avec des liens vers des documents annexes, etc.

---

<sup>4</sup> Les principes du projet Mangrove (<http://www.cs.washington.edu/research/semweb/>) nous paraissent aller dans ce sens : l'objectif est de fournir aux utilisateurs d'un intranet un ensemble de services « sémantiquement pertinents » (agenda, publication, courriel, etc.) dont la valeur ajoutée est immédiatement perceptible, et donc incite à annoter.

Le degré de formalisation de l'ontologie, son étendue, et les possibilités d'inférences liées déterminent tout à la fois le degré d'expertise que son utilisateur doit acquérir pour annoter, la complexité possible de l'interface d'annotation des documents, et les raffinements possibles d'utilisation des méta-données.

Il ne fait pas de doute que les recherches actuelles vont être poursuivies, sans doute dans toutes les directions à la fois, étant donnée l'effervescence actuelle autour du Web sémantique. Cependant, une certaine attention de recherche doit être portée dans quelques directions, moins étudiées mais à notre sens fondamentales et qui, pour beaucoup, devraient impliquer des équipes pluridisciplinaires (Informatique / SHS) :

- Usage actuel (hors du Web sémantique) des méta-données / annotations par rapport aux différentes tâches d'utilisation d'un système d'information documentaire, en lien avec la plus ou moins grande rigueur formelle des schémas de description utilisés. Quels sont les systèmes (d'ajout de méta-données / annotations à des documents) qui fonctionnent actuellement, et pourquoi ? De nombreux travaux ont été menés sur les annotations et les méta-données dans des domaines qui ne relèvent pas du Web sémantique : il serait coûteux de les ignorer.
- Interfaces d'écriture (plus ou moins assistées) et de visualisation de méta-données sur / dans des documents (beaucoup de progrès restent à accomplir). Doit-on annoter avec une ontologie pré-existante, ou bien construire une ontologie adaptée au cours de l'annotation ? Quelle est la sémantique de la relation d'annotation d'une donnée par un concept issu d'une ontologie [1] : que veut-on exactement décrire ? (le monde ? le document ? une utilisation du document<sup>5</sup> ?) Que dire de la notion d'auto-indexation liée à des outils de création de documents Web et ajout de méta-données : quelle en est la signification, comment peut-elle être exploitée ? Peut-on faire confiance à un utilisateur pour bien décrire ses propres documents (à comparer par exemple avec la fonction d'un documentaliste) ?
- Droit, cycle de vie et annotations : diffusion des méta-données, responsabilité, vie privée. Une méta-donnée peut être publiée dans le cadre d'une tâche, ce qui suppose la maîtrise de l'utilisation qui peut en être faite par son auteur. Or des usages non prévus peuvent surgir. Etude de la notion de « validité » des méta-données, par rapport à l'évolution des documents, mais aussi au niveau temporel : cycle de vie, qualité et validation des méta-données / annotations.

---

<sup>5</sup> Voir par exemple [2] où il est proposé d'annoter des ressources par une représentation de leur utilisation, dans le but d'utiliser ces annotations pour la réutilisation des ressources.

Si le travail académique interdisciplinaire autour du Web sémantique se poursuit dans de multiples directions, remarquons que nombre d'outils collectifs d'annotation / commentaires de pages web voient le jour<sup>6</sup>, et que ceux-ci ne se revendiquent pas spécialement du Web sémantique (en particulier, la notion d'ontologie en semble absente). Il est encore trop tôt pour juger de l'acception massive de ces outils ou de leur stabilisation socio-technique. Cependant, leur utilisation des technologies du web et le fait qu'ils participent du développement de l'innovation sur le web (particulièrement le web dit « social » : wiki, blogs, etc.) font qu'ils ne sauraient être négligés par le monde académique s'intéressant aux outils d'annotation / méta-données.

Terminons en remarquant que la question fondamentale qui nous semble posée à long terme est celle du statut des méta-données dans le cadre de l'hypertextualité, liée à la définition, du codage et de la structuration des documents et de l'information sur le Web. En effet, ajouter une méta-données à un document consiste à lui ajouter une structure ; coder un document textuel dans un arbre XML consiste à en définir les éléments structurels essentiels, qui en annotent le contenu textuel ; générer enfin un document adaptatif à partir de multiples sources peut consister à mettre ensemble des éléments de contenu et de structure provenant à la fois de données (documents) et de méta-données pour créer une nouvelle structure, dont le statut est encore peu clair et nécessitera sans doute des travaux théoriques poussés. Les récents travaux du collectif R.T. Pédaque (voir [9] et le numéro spécial de la revue I3 [10]) nous semblent aller dans ce sens.

## 5. REFERENCES

- [1] S. Bechhofer, L. Carr et al. The Semantics of Semantic Annotation. *First International Conference on Ontologies, Databases, and Applications of Semantics for Large Scale Information Systems*. 2159: 1151-1167, Irvine, California, 2002.
- [2] P.-A. Champin et Y. Prié (2003) Musette: uses-based annotation for the Semantic Web, *chapter in Annotation for the Semantic Web*, S. Handschuh (ed.), IOS Press, Amsterdam, pp. 180-190.
- [3] L. Denoue et L. Vignollet. Yawas : un outils d'annotation pour les navigateurs du web. *IHM'99*. Montpellier, France, 22-26 Novembre, 1999.

---

<sup>6</sup> Voir par exemple PageSeeder (<http://www.pageseeder.com/>), Gibeo (<http://gibeo.net/>), Zezame (<http://www.zezame.com/>), Wikalong (<http://wikalong.phunnel.org/wiki/>), StumbleUpon (<http://www.stumbleupon.com/>), etc.

- [4] E. Desmontils, C. Jacquin, et L. Simon, "Annotations sur le Web : Types et Architectures", *Journées Francophones de la Toile (JFT'2003)*, Tours, juin-juillet 2003, pp. 247-256.
- [5] S. Handschuh et S. Staab. Authoring and Annotation of Web Pages in CREAM. *Proceedings of the 11th International World Wide Web Conference*. 462-473. Honolulu, Hawaii, USA, 2002.
- [6] S. Handschuh et S. Staab (Eds).. Annotation for the semantic web. Vol. 96 of *Frontiers in Artificial Intelligence and Applications*, IOS Press. 2003.
- [7] C. C. Marshall. Toward an ecology of hypertext annotation. *Hypertext'98*. New York, ACM Press: 40-49, 1998.
- [8] M. Nilsson, M. Palmer, et al. Semantic Web Metadata for e-Learning - Some Architectural Guidelines. *11th World Wide Web Conference (WWW2002)*. Hawaii, USA. 2002.
- [9] R. T. Pédauque. Document : forme, signe et médium, les re-formulations du numérique. Working paper RTP Document, juillet 2003. [http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/11/index\\_fr.html](http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/05/11/index_fr.html)
- [10] J.-M. Salaun et J. Charlet (Eds) *Revue 13, numéro thématique sur le document numérique*. Vol 4, n°1, juin 2004.
- [11] K.P. Yee, The CritLink Mediator, 1998. <<http://www.crit.org/critlink.html>>