



# Sémantique des folksonomies: structuration collaborative et assistée

Freddy Limpens, Fabien Gandon, Michel Buffa

► **To cite this version:**

Freddy Limpens, Fabien Gandon, Michel Buffa. Sémantique des folksonomies: structuration collaborative et assistée. Ingénierie des Connaissances, May 2009, Hammamet, Tunisia. pp.37-48, 2009. <inria-00378223>

**HAL Id: inria-00378223**

**<https://hal.inria.fr/inria-00378223>**

Submitted on 23 Apr 2009

**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.

# Sémantique des folksonomies: structuration collaborative et assistée

Freddy Limpens<sup>1</sup>, Fabien Gandon<sup>1</sup>, Michel Buffa<sup>2</sup>

<sup>1</sup> Edelweiss, INRIA Sophia-Antipolis, France, 2004 route des lucioles - BP 93,  
FR-06902 Sophia-Antipolis Cedex

{freddy.limpens, fabien.gandon}@sophia.inria.fr

<sup>2</sup> KEWI, Laboratoire I3S, Université de Nice, France

buffa@unice.fr

**Résumé** : L'essor du tagging et des folksonomies pour l'organisation des ressources partagées au sein du Web social et collaboratif constitue une opportunité pour l'acquisition des connaissances par ceux-là même qui les manipulent. Cependant l'absence de liens sémantiques entre les tags, ou la variabilité d'écriture de certains tags appauvrissent les potentiels de navigation et de recherche d'information. Pour remédier à ces limitations, nous proposons d'exploiter l'interaction entre les utilisateurs et les systèmes à base de folksonomies pour valider ou invalider des traitements automatiques effectués sur les tags. Ces opérations se basent sur notre modèle pour l'assistance à la structuration des folksonomies qui autorise des vues conflictuelles portant sur les liens entre les tags, tout en permettant aux concepteurs des systèmes d'exploiter la diversité de ces descriptions sémantiques afin d'offrir des fonctionnalités de navigation enrichies.

**Mots-clés** : Folksonomies, Ontologies, Partage de Connaissances

## 1 Introduction

Le *social tagging* s'est récemment imposé dans le paysage du web social et collaboratif (Web 2.0) comme support à l'organisation de ressources partagées en permettant aux utilisateurs de catégoriser ces ressources, simplement en leur associant des mots clefs, appelés tags. Les folksonomies constituent le résultat de la collecte de tags ainsi créés, c'est à dire associés à des ressources par des utilisateurs. L'exploitation des folksonomies pour la recherche d'informations et de ressources pose néanmoins quelques problèmes. La variabilité d'écriture entre certains tags équivalents (comme "électricité" et "electricite"), ou l'absence de liens sémantiques entre les tags sont pénalisantes lors d'une recherche de ressources par tags. De nombreux travaux de recherche, recensés dans une précédente contribution (Limpens *et al.*, 2008), tentent de dépasser les limitations des folksonomies en les rapprochant de représentations sémantiquement struc-

turées. Ainsi, certains tentent de constituer des “ontologie légères”<sup>1</sup> à partir de folksonomies (Mika, 2005), ou d’assister les cycles de vies des ontologies en les nourrissant des notions extraites des folksonomies (Passant, 2007). D’autres approches proposent d’impliquer les usagers directement dans la construction d’ontologies légères basées sur les formalismes du Web Sémantique (Braun *et al.*, 2007), ou sur le modèle HyperTopic du Web Socio-Sémantique (Cahier *et al.*, 2007).

Notre contribution se concentre sur les systèmes à base de folksonomies outillant le partage de connaissances au sein de groupes de personnes appartenant à des réseaux d’intérêts communs, et regroupés autour de l’usage d’une même plateforme. Le type de système que nous envisageons permet à ses utilisateurs de contribuer au partage, au commentaire, à l’indexation, et à l’élaboration de documents de natures diverses (photos, bookmarks, pages de wiki, etc.). En tant que *ressource* le bookmark a un statut particulier, car il constitue à la fois un support pour documenter les traces de lectures, mais également, et comme par effet secondaire dès qu’il est partagé et associé à des tags, une opportunité pour l’indexation et la (multi)catégorisation collaborative. A ce titre, nous pensons que les pratiques de bookmarking social du Web 2.0 peuvent être adaptées à l’échelle plus restreinte des organisations et des communautés d’intérêts. Le projet ANR ISICIL auquel nous participons s’intéresse notamment à une adaptation de ces outils et usages pour les appliquées à la veille technologique et économique.

Dans cet article nous proposons des méthodes pour constituer des ontologies “légères” qui peuvent être exploitées, par exemple, pour suggérer des termes “sémantiquement” proches lors d’une recherche de documents par tags, ou encore pour enrichir les résultats d’une requête par les variantes orthographiques d’un même tag, ou *spelling variant*, comme “écologie” et “ecologie”. Pour atteindre cet objectif, nous proposons d’allier des traitements automatiques sur les folksonomies et l’expertise des utilisateurs en leur proposant, d’une part, de valider ou invalider les résultats de ces traitements, et d’autre part, de suggérer certaines propriétés sémantiques entre les tags à travers des fonctionnalités simples et ergonomiques de l’interface. Ce système est basé sur notre modèle qui, plutôt que de répercuter et représenter explicitement les traitements automatiques des tags, permet, dans un premier temps, de recueillir les résultats de ces traitements ainsi que les opérations de validation des utilisateurs. L’exploitation de ces résultats est alors repoussée aux étapes ultérieures, par exemple lors du tri ou du filtrage des réponses à une requête.

Notre article est organisé de la manière suivante. Dans la section 2, nous présentons les principales méthodes proposées pour établir des liens sémantiques entre les tags d’une folksonomie, avant de détailler notre méthode pour l’enrichissement sémantique des folksonomies. Nous présentons ensuite dans la section 3 l’implantation de cette méthode dans un système de gestion et d’exploration de bookmarks tagués et partagés. Puis, en section 4, nous positionnerons et discuterons les apports de notre approche, avant de conclure en section 5.

---

1. Gandon (2008) définit les ontologies légère comme étant des “ontologies qui ne comportent typiquement pas ou peu de définitions formelles et qui se focalisent souvent sur la représentation de hiérarchies de types ne nécessitant pas des langages très expressifs (ex : RDFS)”

## 2 Présentation de notre approche

### 2.1 Scénario d'application

Ce travail trouve un terrain d'application dans le cadre d'un partenariat avec l'Ademe (Agence pour l'Environnement et la Maîtrise de l'Energie) où nous cherchons à évaluer la validité des usages liés aux outils du Web 2.0 dans le contexte d'une organisation professionnelle. L'un des scénarios d'application envisagés ici est l'assistance aux pratiques de veille, et plus particulièrement la recherche d'information et le partage de ressources au sein d'un groupe d'experts. Nous souhaitons ainsi promouvoir l'usage du *bookmarking* et du *tagging social* des ressources internes et partagées, ainsi qu'une intégration, dans les tâches quotidiennes des usagers, de l'organisation sémantique des folksonomies.

### 2.2 Traitements sur les folksonomies

Une des limitations couramment reconnue aux folksonomies (Mathes, 2004) est la variabilité d'écriture des tags supposés équivalents comme "écologie" et "ecologie". Une solution possible pour traiter ce problème consiste à mesurer la distance d'édition entre les tags (par exemple de type distance de Levenshtein (1966)), et au delà d'un certain seuil, de considérer ces tags équivalents. Specia & Motta (2007) ont appliqué cette méthode sur un extrait de la folksonomie de delicious.com, et exploité des bases de connaissances externes (Wordnet) et quelques règles simples pour sélectionner le libellé du tag le plus représentatif de ses variantes orthographiques.

Un autre type de traitement des folksonomies consiste à mesurer la distance de "similarité" entre tags en se basant sur les liens entre les tags, les ressources, et les utilisateurs (Mika, 2005). (Cattuto *et al.*, 2008) distinguent sur ce point différents types de mesures de similarité : les mesures basées sur une fréquence de cooccurrence "simple" de deux tags sur une même ressource, ou les mesures distributionnelles, qui prennent en compte trois types de contextes d'association des tags. Chaque contexte correspond à un espace vectoriel prélevé dans l'espace vectoriel global de la folksonomie. Les mesures distributionnelles prennent en compte l'association des tags : (1) *via* leur usage par un même utilisateur (contexte utilisateur-tag), ou (2) *via* leur usage pour une même ressource (contexte ressource-tag), ou (3) *via* leur associations communes avec d'autres tags (contexte tag-tag).

Afin de caractériser en termes sémantiques ces différentes mesures de similarités entre tags, Cattuto *et al.* proposent d'exploiter la structure hiérarchique de Wordnet (Fellbaum, 1998) pour les tags dont le libellé est présent dans cette base lexicale. L'issue de cette expérience montre que les tags associés *via* des mesures de cooccurrences simples tendent à entretenir des relations de subsomption, alors que les tags associés *via* une mesure distributionnelle de similarité dans le contexte tag-tag tendent à se situer au même niveau hiérarchique, soit partageant le même parent, soit le même grand-parent. Cattuto *et al.* expliquent que l'association des tags *via* leur cooccurrence sur une même ressource renvoie à leur utilisation simultanée dans le même acte de tagging où l'utilisateur a tendance à couvrir différents niveaux de généralité. Par exemple, les tags "java" et "programming", ou encore "tobuy" et "shopping" sont fréquemment utilisés

simultanément, et on peut supposer que, du point de vue du “tagueur”, ces tags ont des niveaux différents de généralité. Le lien mesuré par la mesure distributionnelle dans le contexte tag-tag associe des tags ayant des schémas de cooccurrence similaires mais qui ne sont que peu ou pas utilisés simultanément. Ce cas de figure correspond par exemple aux tags “tobuy” et “whishlist” qui ne sont pas utilisés simultanément mais plutôt conjointement avec le tag “shopping”.

La principale limite à l'utilisation de Wordnet comme base de connaissance est que cette ressource termino-ontologique inclue peu de termes spécifiques à un domaine, alors qu'ils sont fréquents dans les folksonomies. Des ressources plus spécifiques à un domaine pourrait donc permettre d'élargir la portée de la validation sémantique des liens de similarités entre certains tags. Cependant la rareté de telles ressources, et la limite de leur couverture d'un domaine repousse toujours plus loin le problème. L'expertise des utilisateurs d'un système semble en définitive la plus adaptée, mais aussi la plus complexe à exploiter si on cherche autant que possible, afin d'éviter toute surcharge cognitive, à limiter l'effort de contribution nécessaire à la formalisation de cette expertise.

## 2.3 Réification des assertions sémantiques sur les tags

L'objectif de notre modèle est de permettre la description des relations sémantiques qui peuvent exister entre des tags, tout en prenant en compte le caractère discutable des assertions portant sur ces relations sémantiques, et ceci autant lorsqu'elles sont le fruit d'un processus automatique que de l'action d'un utilisateur. Ainsi, chaque proposition, validation, ou invalidation de relation sémantique devient un événement dont le système garde une trace. A cette fin nous proposons un schéma RDF/s qui décrit les notions d'assertions et de relations sémantiques en spécifiant les liens qu'elles entretiennent avec d'autres notions issues notamment du modèle RDF de réification des assertions<sup>2</sup>.

Dans notre modèle (voir figure 1), une assertion portant sur la relation sémantique entre deux tags d'une folksonomie est représentée par une classe RDF/s (`TagSemanticStatement`) reliée (par la propriété `hasSemanticRelation`) à une autre classe décrivant la relation sémantique en question (`SemanticRelation` et ses sous-types). De plus, un utilisateur (`sioc:User`<sup>3</sup>, qui peut être aussi un agent automatique) agit sur une assertion sémantique qu'il peut avoir proposée (`hasProposed`), approuvée (`hasApproved`) ou rejetée (`hasRejected`); une assertion sémantique hérite des propriétés de la classe `rdf:Statement` et a donc un sujet (`tag_subject` sous-type de `rdf:subject`) et un objet (`tag_object` sous-type de `rdf:object`). La notion de relation sémantique permet de spécifier, a minima, qu'il existe une relation sémantique entre deux tags, relation qui est spécifiée par ses sous-types dont les significations sont inspirées des propriétés de l'ontologie SKOS<sup>4</sup> : “plus particulier” (`Narrower`); “plus général” (`Broader`); “sémantiquement relié” (`Related`), qui peut être précisée par différents types de mesures de similarités; et enfin “variations

---

2. voir <http://www.w3.org/TR/rdf-mt/#Reif>

3. voir <http://rdfs.org/sioc/spec/>

4. Simple Knowledge Organisation System, <http://www.w3.org/2004/02/skos/>

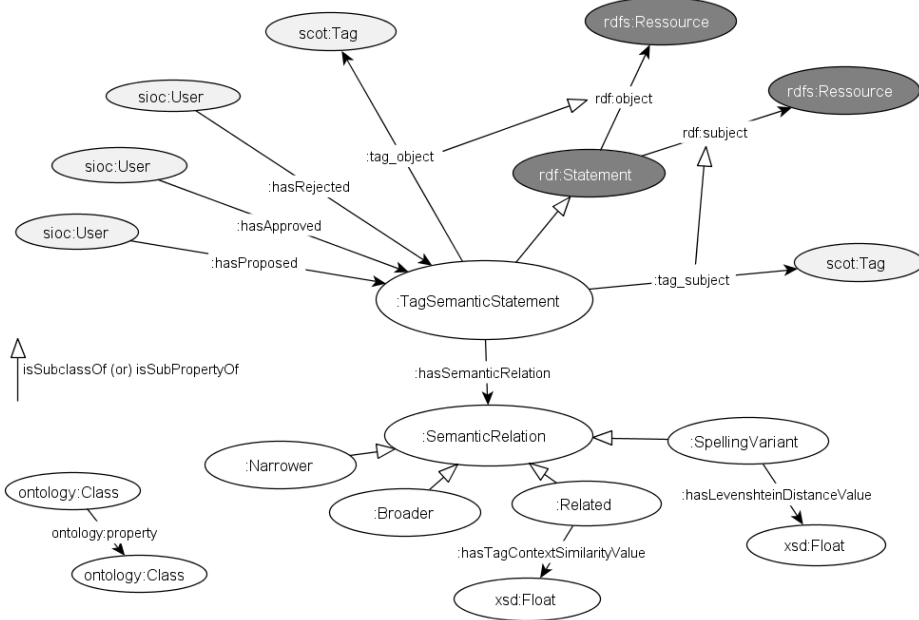


FIGURE 1 – Réification de la notion de relation sémantique

orthographiques” (*SpellingVariant*), qui peut par exemple être spécifiée par la distance de Levenshtein entre les deux tags.

Notre modèle inclue des ontologies déjà existantes comme SIOC (Bojars *et al.*, 2008), ou SCOT (Kim *et al.*, 2007). Le modèle SIOC permet de dissocier la notion de personne de celle d'utilisateur, et finalement d'usage propre, car une personne peut avoir autant d'instances de *sioc:User* que de comptes dans divers services. De plus, le rattachement de chaque instance *sioc:User* à un profil FOAF (Brickley & Miller, 2004) permet à toute personne de réunir et de gérer en un seul point d'accès toute son activité en ligne. L'ontologie SCOT réifie également la notion de tagging (en se rattachant à la Tag-Ontology de Newman *et al.* (2005)) en liant tout évènement de tagging à un compte *foaf:OnlineAccount*, lui même caractérisé par un type précis de plateforme web (blog, forum, etc.). Ainsi, la formalisation de l'acte de tagging nous semble conserver la nature duale du tag, soulignée par Monnin (2009)<sup>5</sup>, qui est à la fois chaîne de caractères interprétable (ici capturé comme label de l'instance de la classe *scot:Tag*), mais aussi "étiquette matérielle" associée à la ressource taguée dans le cadre d'un usage précis et fortement conditionné.

5. que nous rejoignons par ailleurs à propos de l'accent qui doit être mis sur le "design informationnel" des plateformes d'échanges de connaissances, en nécessaire complément aux analyses plus spécifiquement sémantiques.

tag1	tag2	Distance de Levenshtein
informatique	information	0.75
geographie	geographique	0.83
déchets	déchet	0.85
industrie	industriel	0.9
développementdurable	développement-durable	0.95

TABLE 1 – Distance de Levenshtein pour certains couples de tags

### 3 Implantation et résultats

#### 3.1 Détecter les variations orthographiques

En suivant l'exemple de Specia & Motta (2007), nous avons utilisé la méthode de Levenshtein<sup>6</sup> pour mesurer la distance d'édition entre deux chaînes de caractères, ceci dans le but de détecter les variations orthographiques de tags supposés équivalents. Le tableau 1 montre les valeurs de la distance de Levenshtein pour une série de tags extraits d'un échantillon des bookmarks d'utilisateurs de delicious.com ayant utilisé au moins deux fois le tag "ademe"<sup>7</sup> (ou ses variantes orthographiques). A la lecture de ce tableau nous voyons qu'il est délicat de trouver une valeur permettant de dire dans tous les cas que deux tags sont équivalents. Une manière de remédier à ces limitations serait d'employer un dictionnaire, ainsi que certaines règles heuristiques pour valider l'équivalence de deux tags dont la mesure de distance d'édition pour passer de l'un à l'autre passe un certain seuil, ou encore de combiner différentes mesures d'éditions. L'idée étant, dans le cadre de cet article, d'illustrer nos idées relatives à la validation de traitements automatiques par les utilisateurs, et dans l'attente du développement de ces améliorations, nous avons choisis pour notre implantation une valeur seuil de la distance de Levenshtein entre deux tags égale à 0,83.

#### 3.2 Détecter les tags "thématiquement proches"

Nous proposons dans cette partie une méthode qui permet de suggérer des tags "thématiquement proches". Nous nous appuyons dans ce sens sur les résultats de l'étude de Cattuto *et al.* (2008) qui suggèrent dans ce cas l'utilisation d'une mesure distributionnelle de similarité basée sur le contexte tag-tag, par contraste avec les autres mesures distributionnelles ou les mesures basées sur la simple cooccurrence qui ont tendance à refléter des liens de types hiérarchiques. Cette mesure consiste tout d'abord, pour deux tags  $t_1$  et  $t_2$ , à calculer leurs vecteurs associés  $v_1$  et  $v_2$ , où  $v_{ik}$  correspond à la valeur de cooccurrence des tags  $t_i$  et  $t_k$  qui est augmentée d'une unité à chaque fois que les tags  $t_i$  et  $t_k$  sont employés pour le même bookmark. La mesure de similarité entre  $t_1$

6. telle qu'implantée par <http://www.dcs.shef.ac.uk/~sam/simmetrics.html>

7. "ademe" correspond à l'anagramme de Agence De l'Environnement et de la Maîtrise de l'Energie. Notre échantillon se compose des 6054 bookmarks postés par 16 utilisateurs, ayant associé, globalement, 5153 tags distincts à 5969 URL distinctes.

voiture	auto (0.81), automobile (0.83), co2 (0.85), pollution (0.83)
développement	durable (0.88), ecologie (0.8)
construction	habitat (0.95), isolation (0.92), pdf (0.77)
solaire	photovoltaïque (0.74)
réglementation	logement (0.79), thermique (0.82)

TABLE 2 – Pour un tag donné, tags ayant une valeur de similarité dans le contexte tag-tag supérieure à 0.7

et  $t_2$  correspond quant à elle au cosinus de l’angle entre les vecteurs  $v_1$  et  $v_2$ , soit :

$$\cos(v_1, v_2) = \frac{v_1 \cdot v_2}{\|v_1\|_2 \cdot \|v_2\|_2}.$$

Le tableau 2 nous montre une sélection de tags ayant une valeur de similarité (contexte tag-tag) supérieure à 0,7. Pour ces mesures, nous avons prélevé une partie de notre jeu de données en ne conservant que les tags associés aux 100 bookmarks ayant été tagués avec le tag “ademe” (ou une de ses variantes orthographiques)<sup>8</sup>. Nous pouvons observer que les liens inférés reflètent bien les relations thématiques liées au domaine de l’environnement, hormis pour le tag “pdf” qui est associé au tag “construction”. Ceci peut s’expliquer par le fait que le tag “pdf” ait pu être associé souvent aux autres tags liés au tag “construction”, simplement car les documents tagués étaient au format .pdf.

### 3.3 Intégration dans un système de gestions de *bookmarks*

Le système que nous proposons pour illustrer notre propos est un système de navigation au sein d’une base de *bookmarks* extraits de delicious.com (le jeu de données utilisé ici est le même que celui décrit à la section 3.1). Dans notre modèle, nous avons formalisé la notion de bookmark à l’aide de la classe `Bookmark` qui est une sous-classe de la classe `sioc:Item`, faisant du bookmark un document au même titre qu’un billet de blog (`sioc:Post`). De plus, les propriétés `scot:tagOf` et `scot:hasTag` relient le bookmark à un ou plusieurs tags, et la propriété `sioc:about` le relie à une ressource (en tant que `rdf:Resource`), ce qui rend compte de l’indexation faite *via* les bookmarks et permet de retrouver les tags associés à une ressource à l’aide d’une simple requête SPARQL<sup>9</sup>.

Notre système s’appuie également sur notre modèle de réification des relations sémantiques et se compose d’agents automatiques effectuant en tâche de fond des traitements sur les folksonomies, et d’une interface d’exploration de la base de bookmarks. L’imprécision et le caractère discutable des traitements automatiques décrits ci-dessus (cf. sections 3.1 et 3.2) rendent délicate leur application systématique. Nous proposons donc de donner la possibilité aux utilisateurs du système de contribuer à la validation ou l’invalidation des relations sémantiques automatiquement suggérées entre les tags qu’ils manipulent. La figure 2 montre, lors de la recherche de bookmarks par tag, un exemple de fonctionnalité sémantique suggérée par l’interface qui propose une liste

8. cet extrait du jeu de données se compose donc des 100 bookmarks des 75 utilisateurs ayant associé, globalement, 221 tags distincts à 107 URL distinctes.

9. SPARQL Query Language for RDF : <http://www.w3.org/TR/rdf-sparql-query/>



de ressources associées à un tag et ses variations orthographiques (*spelling variant*) calculés grâce à la distance de Levenshtein (la valeur seuil utilisée est 0,83). La fonctionnalité suggérée en question consiste à retirer un des termes de la liste des termes équivalents en cliquant sur la croix rouge encerclée située à côté de chaque terme.

Le recours à la fonctionnalité sémantique proposée reste optionnelle, car l'utilisateur est libre d'interpréter les résultats et peut tout à fait conserver les termes suggérés ("industrial" et "industriel" pour "industrie"). Si toutefois l'utilisateur désire retirer l'un de ces termes, notre modèle laisse le choix aux concepteurs du système d'appliquer cette assertion sémantique (le tag "industrial" n'est pas équivalent à "industrie") pour toutes les requêtes futures d'autres utilisateurs, ou d'un certain groupe d'utilisateurs seulement, ou simplement de ce même utilisateur (comportement que ce dernier attendra certainement en toute logique).

Cet exemple montre la capacité de notre modèle à supporter les actions contradictoires ou conflictuelles des utilisateurs. En effet, lorsqu'un utilisateur choisit de retirer le terme "industrial", le système générera une annotation rendant compte de cette action. Cette annotation s'ajoutera à celle rendant compte du lien d'équivalence entre "industrie" et "industrial", sans l'annuler pour autant. Ainsi notre modèle permet que l'action d'un utilisateur n'annule pas systématiquement celle d'un autre lorsque ces deux actions sont conflictuelles. La décision finale revient aux concepteurs du système qui peuvent choisir différentes "politiques" d'applications des actions de validation par les utilisateurs. Plusieurs solutions sont en effet possibles pour gérer les situations conflictuelles : il est possible de (1) rendre visible ces divergences en les organisant en points de vue qui sont explicitement montrés à l'utilisateur, ou (2) d'appliquer ces divergences différemment selon l'appartenance des utilisateurs à des sous-groupes d'intérêt identifiés par ailleurs au sein de l'organisation considérée, ou encore (3) de proposer un système de vote au sein des groupes d'utilisateurs pour sélectionner l'assertion sémantique à conserver.

## 4 Positionnement et discussion

Dans le cadre de la recherche sur le Web social et sémantique, plusieurs applications concrètes ont implanté des fonctionnalités sémantiques pour organiser des contenus partagés. Les concepteurs de Revyu.com (Heath & Motta, 2007) proposent d'exploiter les formalismes du Web sémantique afin de faciliter l'interopérabilité entre les plateformes de partages de contenus et d'éviter les redondances inutiles. Passant & Laublet (2008) proposent un modèle (MOAT) et des outils qui permettent d'associer les différents sens d'un tag à des documents contenant la définition visée, ou à des concepts d'ontologies du Web Sémantique. CartoDD (Cahier *et al.*, 2007) est l'exemple d'un autre type d'approche basée sur les formalismes du Web Socio-Sémantique (Zacklad *et al.*, 2007), et qui propose d'effectuer une cartographie de contenus à l'aide de cartes de thèmes multi points de vue construites collaborativement par les utilisateurs.

Notre approche quant à elle se démarque de celle de Passant & Laublet (2008) en décrivant, dans un premier temps, le sens des tags grâce à des relations sémantiques entre les tags ("plus général" ou "plus particulier") sans pour autant s'interdire, par la suite et de manière indépendante, de relier les concepts ainsi qualifiés à des concepts

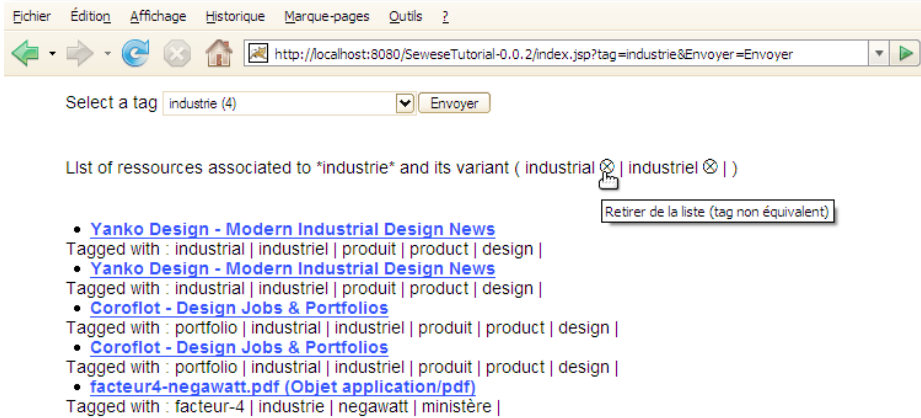


FIGURE 2 – Exemple de fonctionnalité sémantique suggérée par l’interface pour retirer un tag non-équivalent

d’ontologies plus formelles lorsque cela est pertinent pour nos usagers. Notre but est de construire en premier lieu des ontologies légères qui s’apparenteraient aux thésaurus tels que modélisés par le schéma SKOS, et qu’il est toujours possible ensuite de rapprocher d’autres ontologies, formelles ou non, soit en adaptant des techniques d’alignement (Euzenat & Shvaiko, 2007), soit en les mettant en perspectives sous des points de vues différents et explicités dans l’interface, à la manière de (Cahier *et al.*, 2007) qui ont intégré le thésaurus GEMET<sup>10</sup> comme un des points de vue de l’ontologie sémiotique mise en œuvre dans le système CartoDD<sup>11</sup>.

Nous cherchons donc à augmenter les systèmes invitant les usagers à contribuer directement à l’élaboration de vocabulaires partagés en insérant des fonctionnalités d’organisation sémantique dans les interfaces de recherche et de navigation. Ces fonctionnalités consistent à permettre aux utilisateurs de valider ou corriger des suggestions automatiques de termes pertinents pour une recherche d’informations. Elles doivent également rester les moins intrusives possibles afin de ne pas perturber les autres tâches des usagers. Les solutions en cours d’élaboration présentées dans cet article peuvent être vues comme des fonctionnalités complémentaires d’autres outils collaboratifs explicitement dédiés à l’enrichissement sémantique de folksonomies comme par exemple celui développé dans SweetWiki (Buffa *et al.*, 2008).

Notre modèle cherche également à prendre en compte les différents points de vue, à la manière de Cahier *et al.* (2007). En effet, en réifiant la notion de relation sémantique, notre modèle permet de faire de chaque assertion portant sur la sémantique des tags un événement au même titre que le tagging. Même les relations sémantiques entre les tags qui seraient contradictoire (le tag “co2” est plus précis que le tag “polluant” mais également plus précis que le tag “ressource-photosynthèse” par exemple) peuvent être

10. <http://www.eionet.europa.eu/gemet/index.html?langcode=fr>

11. <http://tech-web-n2.utt.fr/dd/?mod=navigation>

recueillies et permettre ainsi de mettre en avant les différents points de vues portant sur une même notion, et de répercuter ces distinctions dans l'ontologie.

L'enrichissement sémantique des folksonomies a également été abordé par Mika (2005) qui propose d'analyser la structure de graphe liée aux folksonomies (*via* les associations ressources/tag/utilisateurs) pour en déduire des liens sémantiques entre les tags. D'autres approches établissent des correspondances entre les tags et des éléments d'ontologies disponibles en ligne sur le Web Sémantique, et étendent les requêtes effectuées sur une folksonomie avec ces éléments d'ontologies (Angeletou *et al.*, 2008). Si nous exploitons le même type de traitements sur les folksonomies que Mika et Angeletou *et al.*, nous cherchons cependant à tirer partie de l'expertise des utilisateurs (à la manière de Tanasescu & Streibel (2007) qui proposent de taguer les tags, où comme Braun *et al.* (2007) qui proposent d'intégrer les approches collaboratives du Web 2.0 dans les processus d'élaboration d'ontologies), ceci afin d'améliorer l'adéquation de ces traitements avec les usages. Par ailleurs, notre approche peut être mise en regard avec celles visant à construire des ontologies de domaines à partir d'une extraction terminologique menée au sein d'un corpus de documents (Aussenac-Gilles *et al.*, 2000). D'un point de vue méthodologique, les tags que nous cherchons à lier sémantiquement à d'autres tags peuvent être vus comme des "candidats-tags" (par analogie aux "candidats-termes"), c'est à dire des syntagmes soumis à la validation d'experts du domaine, avant de devenir éventuellement des concepts ou des relations sémantiques d'une ontologie.

## 5 Conclusion

Notre approche consiste à intégrer les données folksonomiques dans un processus de construction collaborative de représentations des connaissances, et ceci dans le but de fournir des services et des fonctionnalités plus avancées aux systèmes à base de folksonomies. Nous proposons à cet égard d'exploiter des traitements automatiques tout en permettant aux utilisateurs de les valider ou de les invalider. Les deux types de fonctionnalités sémantiques que nous proposons dans cet article sont la reconnaissance de variations orthographiques des tags équivalents et la recherche de tags thématiquement proches<sup>12</sup>. Afin de valider ces inférences automatiques, nous avons montré un exemple de fonctionnalité suggérée par l'interface invitant l'utilisateur à retirer un tag de la liste de tags automatiquement inclus dans la recherche.

Nous avons proposé également un modèle de formalisation des traitements automatiques et des actions de validations par les utilisateurs qui supporte les situations conflictuelles. Nous suggérons ainsi de capturer les assertions, éventuellement divergentes, portant sur la sémantique des tags (et résultantes des traitements automatiques ou de l'action d'utilisateurs), puis de repousser leur traitement au moment de l'exploitation de ces résultats, en fonction des choix des concepteurs du système. Ces derniers peuvent ainsi choisir de montrer les résultats des traitements sémantiques en fonction de l'appartenance à un sous-groupe d'usagers, ou bien en fonction d'une valeur seuil de l'occurrence d'une assertion.

---

12. La première fonctionnalité a été implantée dans notre système de gestion de bookmarks partagés, et la seconde, permettant de suggérer des tags thématiquement proches ("related"), est en cours d'implantation

Nos futurs travaux incluent, outre les tests de terrain avec l'Ademe, la recherche d'autres types de traitements des tags permettant de proposer d'autres fonctionnalités comme la caractérisation plus fine des relations sémantiques entre les tags (relation "plus précis" ou "plus particulier"). D'autres fonctionnalités présentent un intérêt dans la perspective d'outiller l'organisation collaborative des folksonomies, comme la reconnaissance de divergence ou de convergence entre les utilisateurs pour la catégorisation de ressources similaires. Une des applications serait l'assistance à la constitution de groupes d'intérêts qui pourraient, dans le cadre de notre modèle, être utilisés pour personnaliser les interfaces en fonction des annotations sémantiques recueillies lors de l'usage du système. A cet égard, de multiples modalités d'applications des connaissances formalisées ainsi recueillies sont possibles et feront l'objet de recherches futures.

Notre étude s'inscrit également dans la recherche de méthodes pour outiller de manière "dynamique" l'élaboration d'ontologies légères et partagées. A cette fin, nous cherchons à développer des outils permettant aux usagers de saisir au cours de leurs tâches quotidiennes la dimension partagée de leur usage de certains termes. Dans un premier temps, nous avons tenté d'intégrer des fonctionnalités sémantiques aux tâches de recherche d'informations, et dans le cas de notre illustration, plus particulièrement à la recherche au sein d'une base de bookmarks tagués. Nous souhaitons donc étendre notre recherche à l'analyse des usages et des tâches effectuées par les membres d'une communauté ou d'un réseau, dans le but d'identifier d'autres tâches susceptibles d'être autant d'occasions pour l'organisation des connaissances partagées. Une meilleure connaissance des usages permettra également d'accroître les possibilités de personnalisation et de manipulation des résultats donnés par le système, ainsi que son utilisabilité *via* une plus grande transparence des raisonnements appliqués pour obtenir un résultat.

**Remerciements.** Nous remercions l'ANR pour le financement du projet ISICIL ANR-08-CORD-011 qui a permis la production de ces résultats.

## Références

- ANGELETOU S., SABOU M. & MOTTA E. (2008). Semantically enriching folksonomies with flor. In *CISWeb Workshop at Europ. Semantic Web Conf.*
- AUSSENAC-GILLES N., BIÉBOW B. & SZULMAN S. (2000). Corpus analysis for conceptual modelling. In *EKAW - Workshop on Ontologies and Texts.*
- BOJARS U., PASSANT A., CYGANIAK R. & BRESLIN J. (2008). Weaving SIOC into the Web of Linked Data. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China.*
- BRAUN S., SCHMIDT A., WALTER A., NAGYPÁL G. & ZACHARIAS V. (2007). Ontology maturing : a collaborative web 2.0 approach to ontology engineering. In *CKC, volume 273 of CEUR Workshop Proceedings : CEUR-WS.org.*
- BRICKLEY D. & MILLER L. (2004). *FOAF Vocabulary Specification*. Namespace Document 2 Sept 2004, FOAF Project. <http://xmlns.com/foaf/0.1/>.
- BUFFA M., GANDON F., ERETEO G., SANDER P. & FARON C. (2008). SweetWiki : A semantic Wiki. *J. Web Sem.*, **6**(1), 84–97.
- CAHIER J.-P., ZAHER L. & ZACKLAD M. (2007). Information seeking in a "socio-semantic web" application. In *ICPW07 : Proceedings of the 2nd international conference on Pragmatic web*, p. 91–95, New York, NY, USA : ACM.

- CATTUTO C., BENZ D., HOTH O. A. & STUMME G. (2008). Semantic grounding of tag relatedness in social bookmarking systems. *7th International Semantic Web Conference*.
- EUZENAT J. & SHVAIKO P. (2007). *Ontology Matching*. Berlin, Heidelberg : Springer.
- C. FELLBAUM, Ed. (1998). *WordNet An Electronic Lexical Database*. Cambridge, MA ; London : The MIT Press.
- GANDON F. (2008). *Graphes RDF et leur Manipulation pour la Gestion de Connaissances*. Habilitation à diriger des recherches, University of Nice - Sophia Antipolis.
- HEATH T. & MOTTA E. (2007). Revu.com : a Reviewing and Rating Site for the Web of Data. In *ISWC/ASWC*, volume 4825 of *LNCS*, p. 895–902 : Springer.
- KIM H.-L., YANG S.-K., SONG S.-J., BRESLIN J. G. & KIM H.-G. (2007). Tag Mediated Society with SCOT Ontology. In *Semantic Web Challenge, ISWC*.
- LEVENSHTAIN V. I. (1966). Binary codes capable of correcting deletions, insertions and reversals. *Soviet Physics Doklady.*, **10**(8), 707–710.
- LIMPENS F., GANDON F. & BUFFA M. (2008). Rapprocher les ontologies et les folksonomies : un Etat de l'art. *IC*.
- MATHES A. (2004). *Folksonomies - Cooperative Classification and Communication Through Shared Metadata*. Rapport interne, GSLIS, Univ. Illinois Urbana-Champaign.
- MIKA P. (2005). Ontologies are Us : a Unified Model of Social Networks and Semantics. In *ISWC*, volume 3729 of *LNCS*, p. 522–536 : Springer.
- MONNIN A. (2009). Qu'est ce qu'un tag ? entre accès et libellés, l'esquisse d'une caractérisation. In *Ingénierie des Connaissances, Hammamet, Tunisie*.
- NEWMAN R., AYERS D. & RUSSELL S. (2005). Tag Ontology Design. <http://www.holygoat.co.uk/owl/redwood/0.1/tags/>.
- PASSANT A. (2007). Using Ontologies to Strengthen Folksonomies and Enrich Information Retrieval in Weblogs. In *International Conference on Weblogs and Social Media*.
- PASSANT A. & LAUBLET P. (2008). Meaning of a tag : A collaborative approach to bridge the gap between tagging and linked data. In *Proceedings of the WWW 2008 Workshop Linked Data on the Web (LDOW2008), Beijing, China*.
- SPECIA L. & MOTTA E. (2007). Integrating folksonomies with the semantic web. *4th European Semantic Web Conference*.
- TANASESCU V. & STREIBEL O. (2007). ExtremeTagging : Emergent Semantics through the Tagging of Tags. In *ESOE at ISWC*.
- ZACKLAD M., BÉNEL A., CAHIER J., ZAHER L., LEJEUNE C. & ZHOU C. (2007). Hypertopic : une Métasémiotique et un Protocole pour le Web Socio-Sémantique. In *IC*, p. 217–228 : Cépaduès. ISBN 978-2-85428-790-9.