

L'ontologie NiceTag : les tags en tant que graphes nommés

Alexandre Monnin, Freddy Limpens, David Laniado, Fabien Gandon

▶ **To cite this version:**

Alexandre Monnin, Freddy Limpens, David Laniado, Fabien Gandon. L'ontologie NiceTag : les tags en tant que graphes nommés. Atelier Web Social, 10ième Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances, EGC 2010, Jan 2010, Hammamet, Tunisie. pp.12. hal-00452771

HAL Id: hal-00452771
https://paris1.hal.science/hal-00452771

Submitted on 3 Feb 2010

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.

L’ontologie NiceTag : les tags en tant que graphes nommés

Alexandre Monnin*, Freddy Limpens**
David Laniado***, Fabien Gandon**

* EXeCO, Université Paris I Panthéon-Sorbonne
DICEN, Conservatoire National des Arts et Métiers
Alexandre.Monnin@malix.univ-paris1.fr
http://execo.univ-paris1.fr/spip.php?article67
** Edelweiss, INRIA Sophia-Antipolis
{freddy.limpens, fabien.gandon}@sophia.inria.fr
http://www-sop.inria.fr/members/Freddy.Limpens/
http://www-sop.inria.fr/members/Fabien.Gandon/wakka.php?wiki=FabienGandon
*** DEI, Politecnico di Milano
david.laniado@elet.polimi.it

Résumé. Notre analyse part du constat selon lequel les modélisations des tags dont nous disposons actuellement ne prennent pas suffisamment en considération leur richesse et leur diversité. Aussi proposons-nous, pour pallier ce défaut, une ontologie dans laquelle les tags seraient assimilés à des graphes nommés. Ceux-ci sont constitués au minimum d'une ressource reliée à un « signe » qui peut lui-même s'apparenter à n'importe quelle ressource accessible en ligne (un concept d'une ontologie, une image, etc.). Ce modèle entend ainsi fournir une caractérisation suffisamment générale et flexible des tags, et, par voie de conséquence, un cadre susceptible de s'appliquer à tous les tags, quelque soit le modèle sur lequel repose leur description (SCOT, CommonTag, etc.).

1 Introduction

Les tags constituent aujourd'hui un dispositif clef du Web social ainsi qu'un nouveau support d'expression permettant de remplir bien des offices : sélectionner, catégoriser ou classer des contenus, commenter, voter, partager, identifier, etc. Le tagging social et les résultats qu’il engendre (les folksonomies), peuvent être perçus comme des opportunités nouvelles à même d'impliquer les utilisateurs dans une nouvelle forme de commerce vis-à-vis des contenus du Web, libérée des contraintes de l’indexation traditionnelle.

S’il est un trait commun aux modélisations actuelles du tag, c’est bien que ceux-ci y sont représentés comme des instances d'une classe « tag » unique. Cette unicité traduit un rapport univoque aux libellés desdits tags et ce alors même que le libellé d'un tag peut ressortir à une multitude d'emplois, et par conséquent, se voir modéliser de bien des manières différentes en fonction des différents actes de tagging concernés. Qui plus est, la relation entre une ressource taguée et le signe employé pour la taguer est modélisée à l'aide de la seule et

unique propriété « has tag » (SCOT[1]) ou « tagged » (CommonTag[2]). Pourtant, la nature des tags n'est pas univoque ce dont rend compte le modèle CommonTag.

Le but du modèle ici proposé est de founir une modélisation des « actes de tagging » (`TagAction`) qui ne soit pas tributaire d'une interprétation univoque et réductrice de ce que doit être le signe employé en guise de tag. Nous proposons donc, pour décrire les tags de la manière la plus flexible qui soit, de les assimiler en priorité à l’ensemble associant une ressource taguée au signe utilisé pour la taguer, celui-ci pouvant dès lors emprunter différentes formes et conceptualisations (qu'il s'agisse d'une image, d'un litéral, d'un concept issu d'une ontologie, etc.). Ces deux entités sont modélisées à l'aide de la classe `rdfs:Resource` (RDF Vocabulary Description Language), de façon à laisser aux usagers la liberté de mobiliser les modèles du tag ou de la ressource de son choix. Ceci posé, le lien entre la ressource taguée et le signe utilisé pour la taguer est représenté par une propriété et l’assertion obtenue capturée dans un graphe nommé. Sachant que la déclaration des graphes nommés n'est pas nativement supportée en RDF, la décision fut prise de pallier ce manque en intégrant le modèle de Carroll et al. (2005) et la déclarations des sources en RDF/XML proposée par Gandon et al. (2007). Ce choix, rétro-compatible avec les recommandations du W3C, offre en retour la possibilité de mobiliser les différentes initiatives visant à formaliser le tagging (SCOT, CommonTag, etc.) autant que nécessaire et d’établir le lien entre elles de la manière la plus efficace possible – ce qui se traduit, au plan opérationnel, par la possibilité de poser des requêtes portant sur un ensemble hétérogène de modèles. De plus, pour répondre aux problèmes d'ambiguité et d'imprécision des tags, les modèles de tagging actuels, tels que MOAT (Passant et Laublet, 2008), proposent d'associer la signification du tag à chaque acte de tagging. L'ontologie NiceTag, quant à elle, apporte une réponse complémentaire à ce problème en permettant de préciser la relation liant la resource taguée et le tag pour chaque acte de tagging.

Cet article est organisé de la manière suivante. La deuxième section est dévolue à une discussion des motifs nous ayant amenés à proposer un nouveau modèle du tag. Sont détaillés, dans la section suivante, notre modélisation des tags et l'impémentation des graphes nommés qui l'accompagne. La quatrième section présente quant à elle des exemples d'annotations ainsi que des requêtes portant sur des données réelles et impliquant de multiples modélisations. Nous concluons à l'occasion de la cinquième et dernière section.

2 Nature et Usage des Tags

Longtemps la nature des tags est demeurée, au moins partiellement, offusquée. Grâce, cependant, aux travaux menés pour résoudre la "crise d'identité" du Web Sémantique, un arsenal théorique a vu le jour qui nous permet aujourd'hui d'aborder cette question de front. De quoi s'agit-il ? L’augmentation progressive d'un Web de documents en un Web dit sémantique s'avéra bien vite source d'ambiguïté. Effectuée à partir de l’infrastructure du premier, les URIs semblaient d'un coup identifier tout aussi bien des contenus consultables en ligne que des entités absentes du réseau (*en tant que telles*, nous ne parlons évidemment pas ici de leurs descriptions) : personnes, concepts abstraits, fictions, etc.

[1] http://scot-project.org/scot/
[2] www.commontag.org

Une des solutions proposées afin de remédier à ce problème le fut par P.Hayes et H. Halpin. Elle consiste à bien dissocier la *référence*, relation que n'affectent pas les règles techniques de fonctionnement du Web Sémantique, de *l'accès*, traduisant la dimension causale des échanges sur un réseau informationnel tel que le Web. La première doit son fonctionnement à des règles sémantiques, la secondes aux spécifications qui fournissent au Web son assise technologique. Dispositifs matériels autant que sémiotiques, les tags exhibent, *mutatis mutandis*, une dualité similaire à celle des URIs. Dans le sillage des auteurs précédemment cités, nous émettons l'hypothèse selon laquelle il est essentiel d'intégrer les conclusions de ces analyses à tout effort visant à modéliser les tags. Confondre accès et référence serait en effet oublier que le lien symbolique usuel entre mots et choses ne nécessite aucunement de se voir implémenté d'une quelconque manière. Nul besoin d'avoir recours à des moyens d'ordres techniques pour qu'un mot atteigne son objet, aucun artefact n'y pourvoira.

Qu'est-ce en effet qu'un tag, *à première vue*, si ce n'est, pour le dire très grossièrement, l'association, au moyen d'une balise HTML `<a>` et de l'élément `href`, d'un *libellé* qui prend la forme d'une suite quelconque de caractères, d'une images, etc. et d'un lien hypertexte (une URI) ? Contrairement aux vedettes matières ou aux descripteurs dont la sémantique est attachée d'une manière contrainte, soit à un modèle spécifique, soit à un lexique intégralement ordonné par des relations de sens en vue d'éliminer toute ambigüité, le libellé d'un tag est un espace vierge susceptible d'accueillir des entités contrastées, linguistiques ou non, déjouant ainsi toute intelligence globale de la sémantique sous-jacente à son utilisation.

Avec, parfois, l'ajout d'un élément supplémentaire destiné à identifier un lien HTML en tant que tag par l'intermédiaire du microformat `rel="tag"`. Il existe d'ailleurs, à cet égard, une autre manière de caractériser les tags, à notre sens discutable, et qu'illustre ce passage tiré des spécifications du microformat cité à l'instant :

> "By adding rel="tag" to a hyperlink, a page indicates that the destination of that hyperlink is an author-designated "tag" (or keyword/subject) for the current page. Note that a tag may just refer to a major portion of the current page (i.e. a blog post). e.g. by placing this link on a page, tech the author indicates that the page (or some portion of the page) has the tag "tech". The linked page SHOULD exist, and it is the linked page, rather than the link text, that defines the tag. The last path component of the URL is the text of the tag, so fishwould indicate the tag "tech" rather than "fish"[3].

Plusieurs raisons nous conduisent à rejeter ce point de vue :

a) Ce modèle ancre le tagging dans une activité d'emblée communautaire puisque le lien auquel vient se greffer un libellé n'est pas le lien de la ressource qui déclenche l'acte de tagger lui-même mais de la page qui liste l'ensemble des ressources taguées au moyen du « libellé » que l'on retrouve à la fin de l'URI de cette même page. L'activité de l'utilisateur appelé à choisir ses propres libellés (ici « fish ») serait niée si l'on en restait là. Un tag n'étant plus dès lors un signe accolé à une ressource mais une partie d'une URI (ici « tech »).

b) Que le libellé du tag soit ainsi reporté sur l'URI viole un des principes sous-jacent de l'architecture REST du Web sémantique, à savoir le principe d'opacité des URIs[4]. Qui plus est, et l'exemple cité l'illustre, il existe un risque de confusion entre l'ancre textuelle du tag (ici « fish ») et le « libellé » directement intégré à l'URI (ici « tech »).

c) La disparition du lien intention(n)el qui justifie qu'un libellé ait été ajouté à une ressource donnée est éminemment problématique. Des études se sont penchées sur la nature

[3] http://microformats.org/wiki/rel-tag

[4] http://www.w3.org/DesignIssues/Axioms.html#opaque

d'un tel lien afin de mettre en lumière les différentes fonctions du tagging. L'assimilation entre un tag et une URI donnant accès à un ensemble de ressources agrégées par un ou plusieurs individus titulaires d'un compte, ou par la communauté entières des utilisateurs, tend à gommer cette dimension pourtant inhérente à chaque acte de tagging. C'est d'ailleurs là un point que souligne le W3C dans la présentation qu'il donne de ces outils :

> Tagging has emerged as a popular method of categorizing content. Users are allowed to attach arbitrary strings to their data items (for example, blog entries and photographs). While tagging is easy and useful, it often discards a lot of the semantics of the data. A folksonomy tag is typically 2/3 of an RDF triple. The subject is known: e.g., the URL for the flickr image being tagged, or the URL being bookmarked in delicious. The object is known: e.g., http://flickr.com/photos/tags/cats or http://del.icio.us/tag/cats. But the predicate to connect them is often missing. Machine-tags lend themselves to RDF more since they better capture the relationship between the subject and the object. Folksonomy providers are encouraged to capture or infer the semantics around their tags and to leverage semantic web technologies such as RDF and SKOS to publish machine readable versions of their concept schemes. (http://www.w3.org/RDF/FAQ)

Toutefois, l'on aurait tort d'imaginer que ces deux visions s'opposent frontalement. Si l'on ne peut partir de la première pour aboutir à la seconde, l'inverse n'est pas vrai. En partant de ces actes singuliers, qui associent par l'intermédiaire d'un nombre limité de relations typées (cf. section 3.2) *une* ressource à *un* libellé selon des critères d'identité très stricts[5], en les explicitant davantage que cela n'a été le cas jusqu'à maintenant, on solutionne le problème soulevé dans l'extrait précité de la FAQ du W3C lié à l’absence de prédicat. Ceci posé, rien n'empêche ensuite d'affaiblir, volontairement, les critères d'identité du tag. En abandonnant, par exemple, ces relations et les diverses contraintes (de cardinalité ou autre) qui les accompagnent. Ceci afin de lier un libellé non plus simplement à *une* ressource mais à *un ensemble* de ressources, collectées tantôt par le ou les titulaires d'un compte ouvert sur un site de social tagging ou, tout simplement, par l'ensemble des utilisateurs de la communauté. De tels tags « collectifs » se conçoivent alors comme des agrégats de tags individuels. Aussi, au lieu d'aboutir à plusieurs définitions contradictoires les concernant, une solution de continuité émerge par l'application plus ou moins stricte des critères en fonction desquelles les tags sont identifiés.

---

[5] Il faut distinguer ici deux relations irréductibles impliquant des termes différents : la relation entre le libellé et la ressource (que l’on peut modéliser à l’aide de la propriété `irw:refersTo`), et la relation, d'une tout autre nature, découlant des spécifications techniques liée à l'architecture du Web Sémantique, entre la ressource qui déclenche l'acte de tagging et l'URI qui l'identifie et y donne accès (`irw:identifies` et `accesses`). Chaque acte de tagging sur le Web, nommés « tag action » dans notre ontologie, est déclenché par la consultation d'une *ressource en ligne* (les « data item » de la citation précédente) et consiste à lui ajouter un libellé qui lui-même renvoie à une ressource qui peut ou non s'identifier à la précédente. Typiquement, en accédant, via une URI d'un site marchand, au descriptif du roman *La vie devant soi*, paru sous le nom d'Emile Ajar, nous pouvons créer un tag « Romain Gary» et une relation de type « a pour auteur » par exemple qui reliera ainsi le référent de mon tag, non à l'article lui-même mais bel et bien au roman qu'il entend décrire. Bien entendu, dans bien des cas, la ressource à laquelle on accède constituera la référence du tag. *La vie devant soi* réfère, comme nom propre, au livre de R. Gary mais également, dans le contexte quasi-propositionnel du tagging, soit, ici, par ajout de la relation « est à propos de », à la page consultée sur le site marchand (à cet égard les grammairiens et philosophes médiévaux, et le fondateur de la sémiotique moderne lui-même, Charles Sanders Peirce, à leur suite, distinguaient différentes manières de signifier. En particulier la *significatio*, d'ordre lexical, un terme étant pris isolément, et *l'acceptio*, en prise direct, à l'inverse, avec le contexte de l'énonciation).

Voilà, pour résumer les trois principaux aspects où l'ontologie NiceTag innove :

a) en proposant de modéliser le tag au niveau de l'acte singulier de tagging, accompli par un individu (voire une machine lorsque le choix des libellés est automatisé) les tags ainsi définis le sont avec une granularité inédite (cf. section 3.1) ;

b) par la prise en compte des fonctions variées que le tagging remplit, directement ancrée dans les usages et qui les motivent (cf. section 3.2) ;

c) en identifiant les tags non seulement au niveau individuel mais également collectif, par l'assouplissement progressif des critères permettant de les individuer. Ainsi s'effectue le passage du tagging à la folksonomie, des actes singuliers de tagging à leurs agrégats communautaires (avec toutes les nuances intermédiaires). En outillant ce passage progressif de l'un à l'autre, un levier est créé au passage qui permet de répondre aux besoins des diverses communautés amenées à employer le tagging, en mettant tantôt l'accent sur la dimension individuelle de gestion de l'information, tantôt sur la possibilité de favoriser graduellement et à la demande les phénomènes de sérendipité.

3 Modéliser les Tags avec l'ontologie NiceTag

3.1 Les Actions de Tagging en tant que Graphes Nommés

La classe `TagAction` tient donc lieu dans notre modèle, nous l'avons dit, d'équivalent de ce que l'on désigne habituellement sous le vocable « tag ». Elle est modélisée à l'aide d'un graphe nommé contenant des triplets. Ceux-ci ont pour fonction de décrire le lien entre une ressource taguée et un signe. La figure 1 présente le modèle le plus simple de la classe TagAction : une rdfs:resource est liée à une autre rdfs:resource par la propriété « `nt:hasSign` ». Par ce biais, notre ontologie peut être associée à diverses façons de modéliser les ressources taguées et les signes utilisés en guise de tags.

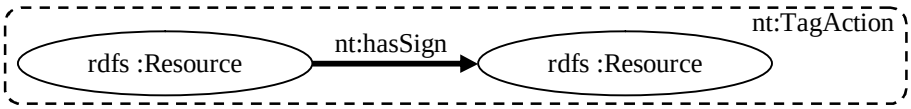


FIG. 1 – *Classe `TagAction` déclarée sous forme de graphe nommé.*

Le graphe nommé « `TagAction` » est déclaré en tant qu'instance de la classe `nt:TagAction`, elle-même une sous-classe de la classe `rdfg:graph` (cf. fig. 2) tirée du modèle de graphe RDF de Carroll et al. (2005). A l'heure actuelle, la syntaxe RDF/XML ne prend pas en charge l'expression des graphs nommés. C'est pourquoi Carroll et *al* ont proposé une nouvelle syntaxe XML pour RDF nommée TriX (Triplets en XML). Gandon *et al.* (2007) ont quant à eux ont proposé une extension rétro-compatible de la syntaxe RDF/XML qui sert à nommer les graphes RDF. Dans la section 3.3 nous décrivons l'implémentation de cette extension afin d'implémenter des instanciations d'actes de tagging.

Pour rendre compte des différentes actions que l'on peut accomplir rien qu'en taguant, nous avons défini plusieurs sous-classes de la classe `TagAction`. Les instances de la classe `TagAction` sont déclarées en tant que graphes nommés résultant d'une action humaine (`ManualTagAction`), pour les différencier de formes plus complexes de tagging telles celles qui impliquent des 'machine tags' (`MachineTagAction`). Agrégées, les actions de tagging peuvent aussi bien revêtir un caractère *collectif* (`CollectiveTagAction`) qu'*individuel* (`IndividualTagAction`). En effet, comme nous l'avons vu, il est souhai-

table de bien distinguer la représentation collective d'un tag, tel que le tag possédant l'adresse `http://delicious.com/tag/improv`, qui renvoie à l'ensemble des signets étiquetés « improv », des représentations individuelles, liées à un compte, de ce même tag, telles que `http://delicious.com/fabien_gandon/improv`, adresse qui pointe vers tous les signets que Fabien Gandon a tagué à l'aide du libellé « improv » (de telles représentations passent par un assouplissement des critères d'identité associés aux tags).

Enfin, la classe `TagAction` est déclarée fille de la classe `sioc:Item` de façon à rendre ainsi compte de la nature partageable des tags. Aussi le tagging peut-il être assimilé à une manière de poster. Ceci nous permet dès lors de décrire l'endroit où les tags sont stockés grâce à la classe `sioc:has_container`, de même que le compte (`sioc:User`) de l'utilisateur (`foaf:Person`) du tag à l'aide de `sioc:has_creator`.

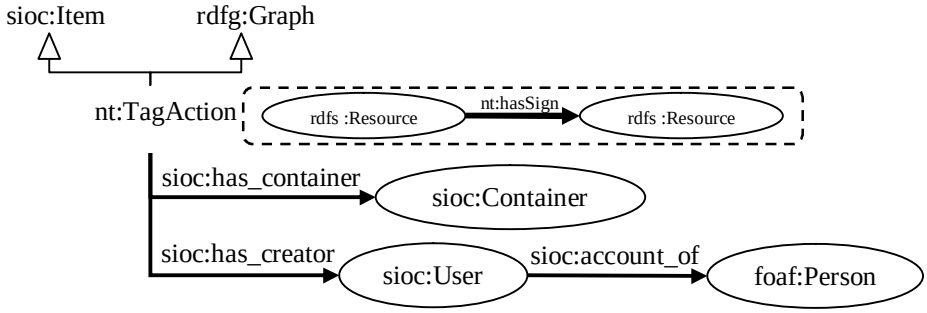


FIG. 2 –*`TagAction` sous-classe de `sioc:Item` et `rdfg:Graph`.*

3.2 Modéliser les usages

Nombre de modélisations actuelles du tag ambitionnent d'associer à celui-ci une signification bien définie ; cette mise en relation est destinée à pallier le problème que représente le fait qu'un terme puisse être doté d'une pluralité de significations selon les contextes ou les communautés qui l'emploient (Passant et Laublet 2008). Seulement, la polysémie est loin d'être la seule source d'ambigüité affectant les tags : une part de leur signification réside en effet dans les types de relations (jusqu'à présent implicites) qui s'établissent entre la ressource et le signe. Par exemple, l'utilisation du tag « blog », l'un des plus populaires sur delicious, pourra sans contredit renvoyer à deux réalités au moins, bien distinctes l'une de l'autre, alors même que le libellé « blog » aura quant à lui été employé conformément une seule et même définition. La ressource peut en effet être *à propos* des blogs ou *être* elle-même un blog (voire les deux à la fois, le tout étant alors de savoir à quoi l'utilisateur fait référence). Qui plus est, certains tags destinés à un usage personnel (idiosyncrasiques) ne font sens que pour un utilisateur précis.

Golder et Huberman (2006) ont identifié pas moins de sept classes de tags à partir des fonctions qu'ils remplissent. Sen *et al.* (2006) ont ramené les classes ainsi mises à jour à trois grandes catégories de tags : factuels, subjectifs et personnels. Des études quantitatives basées sur des applications populaires démontrent qu'une part significative des tags tend à tomber sous les deux dernières catégories (Sen et al. (2006), Al-Khalifa and Davis (2007)). D'autres travaux proposent une classification fonctionnelle fondée sur une première distinction entre tags « liés au sujet » et « non liés au sujet », cette dernière catégorie admettant, à son tour, une subdivision entre tags « affectifs » ou relatifs tantôt à la dimension temporelle, tantôt à

l'accomplissement d'une tâche (Kipp 2008). Les tags liés au sujet sont susceptibles quant à eux d'une caractérisation plus précise qui passe par la distinction entre « liés au contenu » et « liés à la ressource ». Inspirés par ces études, en particulier celle, séminale, de Golder et Huberman, nous avons modélisé les différents usages possibles des tags au moyen de sous-propriétés de la propriété `nicetag:hasSign` (cf. Fig 3.)

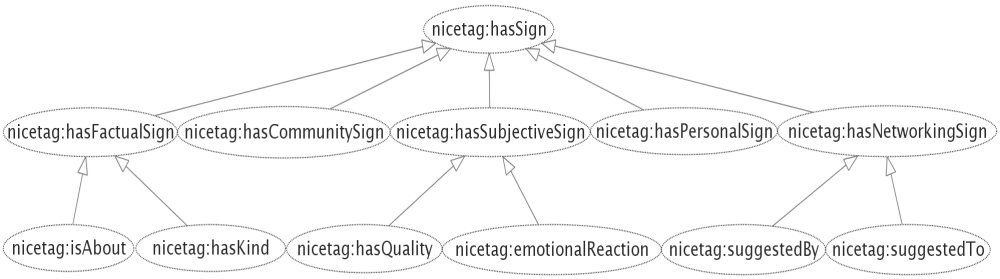


FIG. 3 – *Sous-propriétés de* `nicetag:hasSign`.

La première de ces relations entre signe et ressource est la relation `isAbout`, qui rend compte, sans doute, de l'usage le plus commun du tag, visant à identifier un sujet (le fait d'être *à propos de*). Seconde sous propriété de `hasSign`, `hasKind`, destinée à couvrir l'ensemble des cas où un tag est utilisé pour identifier et caractériser une ressource (forum, vidéo, etc.). La propriété `hasQuality` associe la ressource à un adjectif ou n'importe quel signe exprimant une qualité (joli, idiot, etc.). L'expression des émotions suscitées par une ressource se fait au moyen de la propriété `emotionalReaction` ; typiquement, il s'agira d'exclamations ou d'émoticônes (« wow ! », « ^^ », « :-) »). `hasPersonalSign` couvre l'ensemble des cas où un tag ne fait sens que pour son seul créateur. Ceci inclut tant les classes touchant à l'organisation des tâches de Golder et Huberman que l'ensemble des expressions indexicales ou « token réflexives », qu'elles le soient explicitement (mes_trucs, ma_thèse) ou non (« thèse » pour *ma* thèse, etc.) ; ces dernières trouvant dès lors leur place sans risque d'être confondues avec le sujet ou l'une des caractéristiques quelconques d'une ressource. Dans le même ordre d'idées, la propriété `hasCommunityTag` fut introduite pour rendre compte des tags à destination d'un public ou d'une communauté particulière. A titre d'exemple, nous avons utilisé le tag « #vocampnice2009 » pour partager des ressources concernant le VoCamp où la présente ontologie fut élaborée sur une pluralité de plateformes du Web social. Restent enfin deux propriétés, `suggestedTo` et `suggestedBy`, pour modéliser les actions touchant au *networking*. Plusieurs plateformes ont implémenté pareilles fonctionnalités, notamment delicious en développant pour ce faire une syntaxe spéciale (le double tag « for:username »).

3.3 Utiliser la déclaration de source RDF/XML pour l'implémentation et l'utilisation des graphes nommés

Une requête portant sur une collection de graphes dans SPARQL peut utiliser mot clef GRAPH ou FROM, utilisé pour faire concorder des motifs recherchés (*patterns*) avec des graphes nommés. Le modèle de données RDF se focalise avant tout sur l'expression de triplets dotés d'un sujet, d'un prédicat et d'un objet, cependant ni lui ni sa syntaxe RDF/XML ne fournissent de mécanisme permettant de spécifier la source de chaque triplet. Pour ce faire, il existe une méthode que propose la soumission membre du W3C « RDF/XML Source

Declaration » (Gandon et al. 2007) et qui consiste à associer aux triplets encodés en RDF/XML une URI spécifiant leur origine. Elle requiert l'emploi d'un unique attribut afin de spécifier la source auxquels des triplets exprimés en RDF/XML sont attachés. L'URI de la source d'un triplet est :

1. l'URI source spécifiée par un attribut `cos:graph` associé à l'élément XML encodant ce triplet, si celui-ci existe ; sinon
2. l'URI source du parent de l'élément (obtenu en appliquant les mêmes règles de manière récursive) ; sinon
3. l'URI de base du document.

La portée d'une déclaration de source s'étend de la balise ouvrante dans lequel il apparaît jusqu'à la balise fermante correspondante, à l'exclusion de la portée de toute déclaration interne. Une telle déclaration de source s'applique à tous les éléments et attributs inclus dans son champ. Si aucune source n'est spécifiée, l'URL du document RDF/XML fait alors office de source par défaut. Une seule source peut être déclarée en tant qu'attribut d'un élément donné.

Le code 1.1 montre comment ceci s'applique à un tag capturé dans un graphe nommé. Les lignes 5 à 8 présentent la déclaration du tag en tant que graphe nommé : `http://mysocialsi.te/tag#7182904`. Les lignes 10 à 13 réutilisent le nom de ce graphe pour qualifier le tag en tant que tag créé manuellement par Fabien Gandon le 7 octobre 2009. L'on pourra, à condition de charger au préalable cet ensemble de données RDF dans un entrepôt adéquat, résoudre des requête SPARQL similaires à celle du code 1.2. La ligne 3 correspondant à une recherche portant sur un graphe nommé et le triplet qu'il contient. La ligne 4 permet de s'assurer que ces tags ont été générés manuellement.

Code 1.1. Un tag sous la forme d'un graphe nommé utilisant la syntaxe RDF/XML

```
1  <rdf:RDF xmlns:dc="http://purl.org /dc/elements/1.1/"
2           xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
3           xmlns:cos="http://www.inria.fr/acacia/corese#">
4   <rdf:Resource rdf:about="http://www.yesand.com/"
5                 cos:graph="http://mysocialsi.te/tag#7182904">
6       <nicetag:isAbout>improvisation</nicetag:isAbout>
7   </rdf:Resource>
8   <nicetag:ManualTagAction
9      rdf:about="http://mysocialsi.te/tag#7182904">
10   <dc:creator>Fabien Gandon</dc:creator>
11   <dc:date>2009-10-07T19:20:30.45+01:00</dc:date>
12  </nicetag:ManualTag>
13 </rdf:RDF>
```

Code 1.2. Requêtes SPARQL recherchant des tags créés manuellement.

```
1  SELECT ?t ?a ?g WHERE {
2     GRAPH ? tag { ?t ?a ?g }
3     ? tag rdf : type nt: ManualTagAction }
```

4 Exemples de tags.

Sachant que notre modèle décrit d'abord et avant tout le lien entre la ressource taguée et le signe, sachant également que nous ne posons aucune contrainte tant sur la nature dudit signe que sur celle de la ressource taguée, nous sommes en mesure, par conséquent,

d'exprimer les tags de multiples manières. La figure 4 présente des exemples d'annotations exprimées au moyen de notre modèle. Les actions de tagging font l'objet d'une déclaration sous forme de graphes nommés ainsi qu'expliqué section 3.3 et sont dépeints à l'aide d'une ellipse en pointillés rouges englobant les triplets qu'elle contient. Chaque ellipse représente ainsi une action de tagging et nous avons adopté un code couleur pour distinguer les différentes ontologies mobilisées dans ce schéma. Chaque action de tagging est susceptible de se voir typée à l'aide des sous-classes de `nicetag:TagAction`. Notre exemple provient de données réelles disponibles sur le Web. Les actions de tagging représentées pourraient être typées au moyen de la propriété `ManualTagAction` dans la mesure où elles correspondent toutes à des cas concrets observés chez des utilisateurs de flickr.com ou delicious.com.

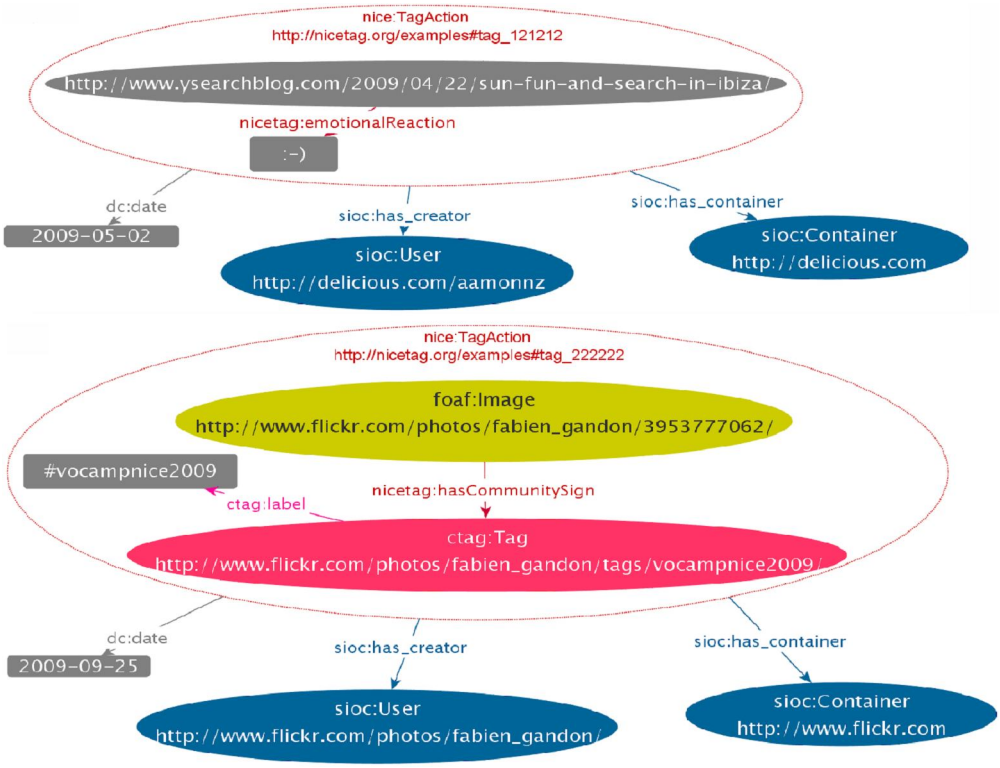


FIG. 4 – *Exemples d'annotations exprimées au moyen de l'ontologie NiceTag.*

Le signe employé pour taguer peut fort bien n'être qu'une suite de caractères telle que « :-) » dans l'exemple de `TagAction`#121212, ou une forme « sémantique » de tagging dans l'exemple de `TagAction`#222222. Rien n'interdit non plus de modéliser les tags à l'aide des instances de la classe `Tag` de l'ontologie CommonTag, de l'ontologie SCOT ou de n'importe quelle ontologie existante ; en un mot, toute `rdfs:Resource` accessible sur le Web. Il convient de noter s'agissant de CommonTag, que la signification du tag « Nice » (donnée par la propriété `ctag:means`) est incluse dans le graphe nommé de l'action de tagging et renvoie à la représentation de la ville de Nice, en France, sur le site geoname.org.

Bien que nous ne contraignions nullement le choix du modèle pour la ressource taguée, nous reconnaissons cependant le caractère essentiel du travail réalisé par Presutti et Halpin

(2009) et présentons des exemples mobilisant la classe `irw:WebResource`, identifiés ici (`irw:identifies`) par l'URL de la page Web de l'événement que constitue le VoCamp qui s'est tenu dans la ville de Nice en 2009. La figure 4 présente des exemples de tagging qui se passent de la classe `irw:Resource`. Dans les deux cas, nous avons recours aux sous-propriétés de `hasSign`. L'action de tagging `#121212` expose par exemple un tag constitué du lien entre une URL et une chaîne de caractère représentant un émoticône (`rdf:Literal` « `:-)` ») mobilisant la propriété « `emotionalReaction` ». Le tag `#222222` présente un cas où une image localisée sur le site flickr.com (modélisée par la classe `foaf:Image`), reçoit un tag `#vocampnice2009` à l'aide de la propriété `hasCommunitySign`. Il s'agit, en l'occurrence, du tag sur lequel se sont accordés les participants du VoCampNice 2009 pour identifier et faire référence à cet événement.

La flexibilité entourant le choix des signes et des ressources permet de récupérer, en une seule requête, tous les actes de tagging exprimés avec notre ontologie, quel que soit le modèle sollicité ou les ressource taguées elles-mêmes. A titre d'exemple, il est possible, en utilisant le moteur RDF CORESE[6], d'écrire les requêtes SPARQL présentées dans le code 1.3. La ligne 2 témoigne de ce que par le biais du mécanisme d'inférence de CORESE, nous sommes en mesure de récupérer tous les types de relations de tagging exprimées à l'aide de `nicetag:hasSign` ou de ses sous-propriétés. La même chose vaut pour tous les types subordonnés de la classe `nicetag:TagAction` (ligne 7). La ligne 3 montre, avec l'assertion OPTIONAL, que notre modèle est capable de récupérer des ressources utilisées pour taguer, fussent-elles typées ou non.

Code 1.3. Requête sur des actions de tagging réparties selon plusieurs modèles.

```
1  SELECT * WHERE {
2   GRAPH ? tagaction {?resource nicetag:hasSign ?sign }
3   OPTIONAL {
4    ?sign rdf:type ?signtype .
5    ?sign rdfs:label ?signlabel .}
6   ?resource rdf:type ?resourcetype .
7   ?tagaction rdf:type nicetag:TagAction }
```

Code 1.4. Requête sur des actions de tagging liées à différents comptes d'utilisateurs.

```
1  SELECT * WHERE {
2    GRAPH ? tagaction {?res nicetag:hasSign ?sign }
3    ?tagaction sioc:has_creator ?user .
4    <http://ns.inria.fr/fabien.gandon/foaf#me> foaf:holdsAccount ?user .
5    ?tagaction rdf:type nicetag:TagAction }
```

Ces exemples présentent également une distinction entre créateurs et containers des actions de tagging. Les utilisateurs sont modélisés en tant qu'instances de la class `sioc:User`. En ajoutant des triplets pour connecter les différents comptes d'une personne (modélisée à l'aide de `foaf:Person`) avec la propriété `foaf:holdsAccount`, il devient loisible de rapporter tous les actes de tagging d'une personne donnée grâce à la requête présentée dans le code 1.4. Le container d'une action de tagging est modélisé au moyen de la classe `sioc:Container` qui permet de poser des requêtes de type « récupérer tous les tags provenant de delicious.com et seulement ces derniers » (cf. code 1.5).

Code 1.5. Requêtes visant à récupérer les actions de tagging de delicious.com

```
1  SELECT * WHERE {
2     GRAPH ?tagaction {?resource nicetag:hasSign ?sign }
```

[6] http://www-sop.inria.fr/edelweiss/wiki/wakka.php?wiki=Corese

```
3     ?resource rdf:type ?resourcetype .
4     ?tagaction rdf:type nicetag:TagAction .
5     ?tagaction sioc:has_container <http://delicious.com > }
```

5 Conclusion.

L'essence du tag telle que la conçoit l'ontologie NiceTag consiste à donner les moyens d'enregistrer la trace d'actions associant une ressource avec un signe grâce à un triplet RDF. Dans l'optique d'intégrer à nos spécifications les fonctions qu'est susceptible de remplir un tag, nous avons créé plusieurs sous-propriétés qui couvrent les divers types de relations qu'entretiennent le tag et la ressource taguée. Cet ensemble est à son tour conçu à la manière d'une instance de la classe `TagAction` et peut en conséquence être enrichi par toutes les propriétés qui lui sont associées, au nombre desquelles figure celle qui spécifie l'utilisateur qui a accompli l'action de taguer, sa date ou son container. Qui plus est, il est possible de définir le genre auquel appartient l'action de tagging (automatique, individuel, collectif…) en choisissant l'une des sous-classes de `TagAction` définies à cet effet. Ce faisant, et en vertu du recours à la déclaration de source pour la syntaxe RDF/XML, qui assigne une URI à une action de tagging, nous obtenons une grande expressivité pour représenter les tags sous une multitude de facettes tout en évitant le fardeau de la réification. Tant les graphes nommés que la déclaration de source pour la syntaxe RDF/XML fournissent une plus-value notable au prix de modifications modestes et rétro-compatibles avec les recommandations officielles relatives au Web Sémantique.

Combiné avec les vocabulaires consacrés au tagging, notre modèle offre un cadre des plus flexibles pour réaliser l'interopérabilité des systèmes de social tagging. Le recours à l'ontologie NiceTag et aux langages de requête SPARQL permet d'agréger et de poser des requêtes sur une variété de sources et de représentations. Quant aux modèles actuels de tagging ils proposent d'enrichir la représentation des actions de tagging en se focalisant sur le signe utilisé pour taguer, qu'il s'agisse d'associer un tag à un concept du Web Sémantique bien définie (avec MOAT ou CommonTag) ou de spécifier les relations morphologiques ou lexicales entre tags (à l'aide de SCOT). L'ontologie NiceTag permet pour sa part de préciser la nature de chaque acte de tagging en précisant la relation entre signe et ressource taguée. Ainsi, devient-il aisé de lever l'ambigüité des tags dès lors qu'un même signe, tout en conservant la même signification, peut être utilisé tant pour spécifier la thématique d'une ressource (`isAbout` « blog ») que son type (`hasKind` « blog »).

Concernant les risques de surcharge cognitive inhérents à notre modèle, nous constatons qu'il est très difficile de naviguer après-coup dans un ensemble de tags dépassant une taille relativement modeste. Les approches (CommonTag, MOAT) offrant de remédier à ce problème en proposant de spécifier la signification des tags (entendus ici comme de simples libellés) nous semblent pour leur part impliquer un effort non négligeable de la part des utilisateurs qui ont ou auront à choisir parmi un vaste ensemble d'entrées possibles. L'alternative que nous proposons nous semble à tous égards plus économique. Elle consiste à choisir parmi un nombre comparativement extrêmement limité de relations pragmatiques, indissociables des usages concrets des tags et du point de vue des utilisateurs. La motivation accompagnant cet effort a toutes les chances d'être plus forte, car tournée d'avantage vers l'usage personnel et circonstancié, que l'action d'accrocher aux tags des « significations » non ambigües - ce qui relève semble relever davantage, disons-le, d'une forme d'altruisme.

Références

Al-Khalifa H. S. et Davis H. C. (2007), Towards better understanding of folksonomic patterns, Hypertext 2007, 163–166.

Carroll J. J., Bizer C., Hayes P. and P. Stickler (2005). Named graphs, provenance and trust. In WWW '05: Proceedings of the 14th international conference on World Wide Web, 613–622, New York, NY, USA: ACM.

Gandon F., Bottolier V., Corby O. et P. Durville (2007), RDF/XML source declaration, w3c member submission. http://www.w3.org/Submission/rdfsource/.

Golder S. A. and Huberman B. A. (2006), Usage patterns of collaborative tagging systems. *J. Inf. Sci.*, 32:2, 198–208.

Halpin H. and Presutti V. (2009), An ontology of resources: Solving the identity crisis. *LNCS*, 5554, 521–534.

Hayes P. J. & Halpin H. (2008), In defense of ambiguity. *Int. J. Sem Web Inf. Sys.*, 4:2, 1–18.

Kim H.-L., Scerri S., Breslin J., Decker S. and H.-G. Kim (2008), The state of the art in tag ontologies: A semantic model for tagging and folksonomies.

Kipp M. E. (2008), @toread and cool: Subjective, affective and associative factors in tagging.

Passant A. et 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.

Sen S., Lam S. K., Rashid A. M., Cosley D., Frankowski D., Osterhouse J., Harper F. M. and J. Riedl (2006), tagging, communities, vocabulary, evolution, 181–190.

Wolff C., Heckner M. and S. Mühlbacher (2008). Tagging tagging. Analysing user keywords in scientific bibliography management systems. *Journal of Digital Informaton*, 9:27.

Summary

Current tag modelling does not fully take into account the rich and diverse nature tags, as signs, can take on. We propose an ontology of tags in which tags are modelled as named graphs. These named graphs are made of a resource linked to a "sign" which can be any resource reachable on the Web (an ontology concept, an image, etc.). The purpose of our model is to be able to describe tags in a very general manner, and as an immediate consequence, to describe tags as modelled by other tag models (SCOT, CommonTag, etc.).