



HAL
open science

Rapprocher les ontologies et les folksonomies pour la gestion des connaissances partagées : un état de l'art.

Freddy Limpens, Fabien Gandon, Michel Buffa

► To cite this version:

Freddy Limpens, Fabien Gandon, Michel Buffa. Rapprocher les ontologies et les folksonomies pour la gestion des connaissances partagées : un état de l'art.. 19es Journées Francophones d'Ingénierie des Connaissances (IC 2008), Jun 2008, Nancy, France. pp.123-134. hal-00416693

HAL Id: hal-00416693

<https://hal.science/hal-00416693>

Submitted on 14 Sep 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.

Rapprocher les ontologies et les folksonomies pour la gestion des connaissances partagées : un état de l'art.

Freddy Limpens¹, Fabien Gandon¹, Michel Buffa²

¹ Edelweiss, INRIA Sophia-Antipolis, France, 2004 route des lucioles - BP 93,
FR-06902 Sophia-Antipolis Cedex
{freddy.limpens, fabien.gandon}@sophia.inria.fr

² KEWI, Laboratoire I3S, Université de Nice, France
buffa@unice.fr

Résumé : Ces dernières années, le “tagging social” s’est imposé au sein du Web 2.0 comme le principal moyen de classification de données en très grand nombre. Néanmoins, l’emploi pour la classification d’ensembles de tags non contrôlés, appelés folksonomies, pose plusieurs problèmes : le problème de l’ambiguïté (un tag dénotant plusieurs concepts), le problème de variations d’écritures (plusieurs tags dénotant un même concept) et le manque d’assistance à l’exploitation de ces structures (notamment pour la recherche ou l’échange d’informations) en l’absence de représentations explicites de ces connaissances partagées. Cet article présente un état de l’art comparatif et une discussion sur les travaux proposant une évolution de ces systèmes à base de folksonomies, en tenant compte, notamment, des travaux existant dans le domaine de la représentation des connaissances, en particulier ceux à base d’ontologies.

Mots-clés : Web social, Folksonomies, Ontologies, Partage de connaissances

1 Introduction

La classification et le partage des ressources au sein du Web posent plusieurs problèmes auxquels prétendent répondre les systèmes de représentation à base d’ontologies ainsi que ceux reposant sur des folksonomies. Le terme folksonomie est un mot valise d’origine anglaise introduit par Vanderwal (2004) et exprimant l’idée d’une classification (“taxonomy”) faite par les usagers (“folks”). Les ontologies sont des structures de représentation d’un champ de connaissance et visent à “spécifier explicitement une conceptualisation” (Gruber, 1993). Plus particulièrement, les ontologies formelles sont des spécifications de la conceptualisation d’un champ de connaissances à l’aide de sémantiques formelles, et sont exploitables par des programmes informatiques afin d’assister le partage de ces connaissances à travers différentes applications. Leur conception relève souvent d’un processus engageant des ontologues, assistés d’experts

du domaine à modéliser. L'obstacle majeur à la généralisation de l'emploi d'ontologies formelles au sein du Web est le coût de leur élaboration, de leur maintenance, et de leur exploitation pour l'annotation des ressources.

Le partage de ressources taguées sur le Web 2.0 s'appelle le "tagging social" et produit des folksonomies. Les folksonomies sont des organisations de ressources issues des usages et leur obtention est non supervisée. Ces folksonomies ouvrent de nombreuses possibilités d'exploitation, pourvu qu'un système de partage efficace des connaissances et des ressources en tire profit. Dans cette perspective se dressent cependant certains obstacles à l'emploi des folksonomies comme représentation des connaissances (Mathes, 2004; Passant, 2007) : (1) l'ambiguïté des tags, car un tag peut désigner plusieurs concepts ; (2) les variations des écritures des tags pour désigner un même concept ; (3) et le manque d'assistance à l'exploitation de ces structures (notamment pour la recherche ou l'échange d'informations) en l'absence de représentations explicites ou formelles de ces connaissances partagées.

Un certain nombre de travaux se sont penchés sur ce problème de l'annotation des documents et du partage des connaissances. Les systèmes qu'ils proposent tentent de réconcilier les modèles de représentations des connaissances à base d'ontologies et ceux à base de folksonomies. Dans une première section nous présentons les approches visant à extraire des liens sémantiques entre les termes des folksonomies. Dans une deuxième section, les contributions présentées s'attachent à soutenir les folksonomies à l'aide d'ontologies. Enfin nous abordons les travaux visant à exploiter les formalismes du Web sémantique pour l'échange de ressources.

2 Émergences de structures dans les folksonomies

Les contributions présentées dans cette partie analysent a posteriori la structure des folksonomies pour obtenir une représentation de la structure sémantique induite par les communautés, comme première étape supposée à l'élaboration d'ontologies.

2.1 Analyse des réseaux sociaux appliquées aux folksonomies

Mika (2005) propose de considérer les folksonomies comme des structures sémantiques émergentes des usages de la communauté. Son objectif est de construire des "ontologies légères" (lightweight ontologies) à partir de l'analyse de folksonomies. Mika propose un modèle "tripartite" des folksonomies où les instances sont des ressources Web associées par un acteur à une liste de concepts (les tags). Des méthodes d'analyse de réseaux sociaux sont ensuite utilisées pour tisser des réseaux reflétant les liens entre les concepts et en déduire des regroupements de termes, ou des relations de subsumption (relation "is-a"). Mika (2005) rapproche les concepts via deux perspectives en tissant (1) le réseau liant les instances (ressources) aux concepts (tags) les caractérisant, et en rapprochant les concepts (tags) ayant le plus d'instances en commun (2) le réseau liant les concepts aux acteurs les utilisant et en rapprochant les concepts ayant le plus d'acteurs en commun. Le premier réseau donne un graphe concepts-instances reflétant la co-occurrence des tags sur les mêmes ressources. Le deuxième donne un graphe

concepts-acteurs reflétant les recouvrements entre communautés d'intérêt. Dans ce dernier cas, une communauté d'intérêt pourrait être représentée par l'ensemble des acteurs ayant utilisé le tag "longitude". Si les communautés "longitude" et "latitude" ont un certain nombre d'acteurs en commun, les tags "longitude" et "latitude" seront considérés comme sémantiquement proches. Le graphe concepts-acteurs reflète de manière plus précise la spécificité du vocabulaire utilisé par la communauté considérée, car le rapprochement de deux termes ne peut se faire que s'ils ont un grand nombre d'utilisateurs en commun. Il permet de surcroît de déduire des relations du type "plus général / plus précis" entre les termes. En effet si l'ensemble correspondant à la communauté "pêche sous-marine" est un sous-ensemble de la communauté liée au tag "activités nautiques", alors le tag "activités nautiques" est considéré comme plus général que le tag "pêche sous-marine".

2.2 Analyse de la dynamique des folksonomies

Halpin *et al.* (2007) analysent la dynamique des folksonomies et recherchent les lois de distribution de la fréquence d'utilisation des tags. La question qu'ils posent est celle de la cohérence des folksonomies et, plus précisément, celle de la stabilité de la distribution des tags pour une ressource donnée au cours du temps. Leur hypothèse est que les tags les plus utilisés pour annoter une ressource demeurent les mêmes, et que la distribution de leur fréquence d'apparition suit une loi de puissance. Ils vérifient cette hypothèse pour les sept à dix tags les plus utilisés pour caractériser des sites populaires sur un service de *bookmarking* social¹. L'analyse montre que la distribution des tags pour les sites les plus populaires devient stable à partir d'un temps quasiment constant. Par ailleurs, les auteurs se sont intéressés à la recherche de liens sémantiques entre les termes de ce vocabulaire à l'aide de graphes de corrélations inter-tags. Chaque noeud de ces graphes représente un tag sous forme d'un cercle dont le diamètre est pondéré par la fréquence d'apparition de ce tag. La longueur des arcs reliant les tags co-occurents est pondérée par leur degré de co-occurrence. Cette visualisation est un point de départ possible pour l'élaboration d'ontologies plus structurées.

2.3 Cluster en correspondance avec des ontologies

Specia & Motta (2007) effectuent un traitement statistique des annotations en regroupant les tags fortement liés entre eux pour former des clusters. La méthode choisie consiste à construire la matrice de co-occurrence des tags sur les mêmes ressources, puis à calculer la distance angulaire entre chaque ligne de cette matrice. Cette distance est exploitée pour constituer des clusters de tags. La deuxième étape consiste à expliciter les liens pouvant exister entre les tags d'un même cluster. Pour ce faire, le système recherche dans des ontologies du Web sémantique la présence de toutes les paires de tags de chaque cluster. En cas de succès, le système recherche les relations possibles existant entre les tags de chaque paire. Une tentative d'automatisation plus complète de cette méthode a par ailleurs été menée par Angeletou *et al.* (2007).

¹<http://del.icio.us>

2.4 Fouille de donnée et folksonomies

D'autres approches reprennent le modèle tripartite des folksonomies de Mika (Mika, 2005) et appliquent des méthodes de fouille de données afin d'extraire des informations sur la structure des folksonomies. Jäschke *et al.* (2008) exploitent les méthodes d'analyse formelle de concepts afin de découvrir les sous ensembles d'utilisateurs de folksonomies partageant la même conceptualisation sur les mêmes ressources. Pour ce faire ils construisent des triples (Ressources, Utilisateurs, Tags) appelés tri-concepts où chaque utilisateur a tagué chaque ressource avec tous les tags. L'extraction des tri-concepts d'une folksonomie constitue, selon les auteurs, une première étape à la construction d'ontologies. Les ontologies sont ainsi vues comme des constructions sociales où un concept est décrit par un ensemble de tags appartenant à un ensemble d'utilisateurs et utilisés pour caractériser un certain type de ressources.

D'autres techniques de fouille de données ont été appliquées par Schmitz *et al.* (2006) afin d'extraire des règles d'association au sein des folksonomies. La première étape est de projeter le modèle tripartite (Ressources, Utilisateurs, Tags) sur une structure à deux dimensions, par exemple en considérant l'ensemble des couples (Utilisateurs, Ressources) associés à un ensemble de tags T_x . A partir de cette projection, un exemple de règle d'association qu'il est possible d'extraire est le suivant : les Utilisateurs associant les tags de l'ensemble T_A à un ensemble de Ressources, associent souvent les tags de l'ensemble T_B à ces mêmes ressources. Ce type de règles pourrait être notamment appliqué dans un système de recommandation de tags.

2.5 Structuration manuelle de folksonomies

Tanasescu & Streibel (2007) proposent pour structurer les folksonomies de réexploiter les ressorts ayant fait le succès des applications du Web 2.0. Ils suggèrent d'étendre les systèmes actuels de tagging social en permettant à tous les utilisateurs de taguer les tags eux mêmes ou les relations pouvant exister entre certains tags. Ces relations peuvent naturellement être exprimées à l'aide de triples, ce qui permet d'intégrer facilement ces connaissances dans un moteur de recherche du Web sémantique. Les auteurs soulèvent le problème de la pertinence des informations fournies par les utilisateurs qu'un système non centralisé ne peut garantir. Afin d'y remédier, la régulation d'un tel système serait assurée par la communauté (cf. WikiPedia), et outillée par la possibilité d'apprécier ou de déprécier un tag, ou encore par des systèmes de "points" gagnés en fonctions de la participation en tant qu'auteur ou même navigateur. Ces procédés visant à inciter les contributions sont propres aux plateformes sociales et pourraient par ailleurs être étendus avec les "jeux sérieux". Les "jeux sérieux" sont des activités présentés sous un jour ludique, mais exploitées à des fins de traitement de données complexes nécessitant de l'intelligence humaine à moindre coût. Un exemple de ce type de jeux est "The ESP Game"² où deux joueurs connectés simultanément doivent se mettre d'accord sur le label (tag) à attribuer à une image qu'ils ont sous les yeux. Les multiples annotations d'images ainsi fournies par les joueurs sont ensuite capitalisées et mises à disposition pour la recherche d'informations, ou la description d'images aux non-voyants.

²<http://www.espgame.org>

2.6 Conclusion partielle

Les résultats des approches visant à extraire la structure sémantique des folksonomies sont encourageants. Mika (2005) montre qu'il est possible d'extraire d'une folksonomie les termes les plus représentatifs d'une communauté ainsi que certaines relations sémantiques entre ces termes. Halpin *et al.* (2007) montrent de leur côté que les folksonomies liées aux sites les plus populaires présentent des distributions stables et peuvent aussi servir de base à la construction d'ontologies. Les méthodes de fouille de données proposées par Schmitz *et al.* (2006) et Jäschke *et al.* (2008) apportent des perspectives sur les folksonomies en vue d'une explicitation plus avancée. Cependant les structures extraites automatiquement souffrent encore de l'ambiguïté des liens établis entre les concepts. Si nous prenons l'exemple des concepts "véhicules, avions, pilotes", il serait possible, à partir des méthodes proposées par Mika, d'inférer un lien entre ces trois termes, puis un lien de subsumption entre "avions" et "véhicules". Mais le lien entre "avions" et "pilotes" n'est pas un lien de subsumption, et le typage de ce lien requiert une expertise du domaine concerné. Reposant sur le dynamisme du Web social, Tanasescu & Streibel (2007) suggèrent quant à eux d'impliquer les utilisateurs dans la qualification des liens entre tags, tandis que Specia & Motta (2007) entreprennent cette tâche en interrogeant les ontologies du Web sémantique.

3 Ontologies comme support des folksonomies

Dans cette partie nous abordons les travaux proposant d'apporter un support formel aux folksonomies, soit en faisant des tags les attributs des concepts d'une ontologie formelle (Passant, 2007; Good *et al.*, 2007) soit en réifiant les tags eux mêmes et en créant une "ontologie de folksonomie" (Gruber, 2005; Buffa *et al.*, 2008; Newman *et al.*, 2005; Kim *et al.*, 2007).

3.1 Guider le tagging à l'aide d'ontologies

De nombreuses solutions visant à intégrer le plus naturellement possible interface de tagging et systèmes à base d'ontologies ont récemment vu le jour. Passant (2007) propose de structurer une folksonomie pendant l'annotation des billets d'un blog d'entreprise par l'intermédiaire d'un système centralisé ; ici un tag est une propriété (has-Tag) d'un concept d'une ontologie contrôlée. L'ontologie subordonne l'annotation en aiguillant le choix du tag : il s'agit en fait pour l'utilisateur de réutiliser un label existant, ou de proposer un nouveau label pour un concept qui existe déjà. Cependant l'utilisateur reste libre de soumettre à l'administrateur un concept et son label (tag) s'il n'existe pas encore. L'intérêt vient de ce que l'ambiguïté des tags par rapport aux concepts qu'ils sont censés désigner est automatiquement levée : si l'utilisateur propose un tag qui peut être relié à deux concepts différents (par exemple "EDF" pour "Électricité De France" ou pour "Équipe De France"), le système propose de choisir parmi les différentes possibilités. Le billet annoté par son auteur sera lui aussi relié au concept sélectionné afin de lever l'ambiguïté des tags au moment de la consultation du blog, en permettant de ne récupérer que les billets liés au concept voulu et non ceux liés à tous les concepts

associés à un tag donné. Suivant une démarche similaire, Good *et al.* (2007) proposent un accès à des terminologies professionnelles aux utilisateurs d'un service de *bookmarking* social, Connotea³. Ainsi les utilisateurs ajoutant des tags à une ressource peuvent choisir parmi une suggestion, faite par le système, de termes extraits d'ontologies. Le système faisant coexister des termes librement choisis et des termes issus de terminologies contrôlées, les auteurs suggèrent d'exploiter l'ensemble de ces annotations pour en déduire des relations existantes entre ces deux types de termes. Dans une perspective de généralisation, il semble que ce type de méthode nécessitent de connaître a priori le domaine de connaissances concerné afin de guider la sélection des ontologies à interroger pour suggérer des termes.

3.2 Construire une ontologie de folksonomie

Dans son article, Gruber (2005) pose qu'il n'y a pas d'opposition entre les ontologies et les folksonomies et propose de construire une "ontologie de folksonomie". La "TagOntology" est un projet de construction d'une ontologie commune dédiée à la formalisation et la conceptualisation de l'acte d'annoter une ressource par un terme (tagging). Ce modèle met en œuvre quatre entités pour caractériser un événement de "tagging" : l'objet tagué, ou la ressource ; le tag, c'est à dire le mot-clé utilisé ; l'utilisateur taguant ; et le domaine au sein duquel le tagging s'inscrit (ce peut être le service utilisé par exemple). A la différence de Passant (2007), pour qui le tag est une simple chaîne de caractères reliée à des concepts d'ontologies formelles, Gruber va plus loin dans la réification des tags et considère chaque tag comme un objet à part entière. Pour contrer les ambiguïtés des tags ou les usages abusifs (*spam*), Gruber proposait déjà de "tagger les tags". Par ce biais, il serait possible d'indiquer que tel tag est synonyme de tel autre tag, ou encore que tel tag est adéquat ou non pour tel objet. Kim *et al.* (2007) proposent quant à eux l'ontologie SCOT⁴ pour représenter les nuages de tags afin de faciliter leur partage (section 4.2). Buffa *et al.* (2008) ont de leur côté appliqué l'idée d'exprimer les tags directement à l'aide des langages du Web sémantique. Les tags sont donc collectés comme autant de nouvelles classes d'ontologie et servent à annoter les pages d'un wiki "sémantique" (section 4.3).

3.3 Conclusion partielle

Dans cette partie nous avons montré que les approches à base d'ontologies peuvent être appliquées aux folksonomies, apportant une richesse supplémentaire illustrée par des exemples concrets d'implémentation. Passant (2007) fait de chaque tag une propriété rattachée à des concepts dans l'ontologie d'une entreprise. Cette approche permet, au cours de l'utilisation, de lever l'ambiguïté des tags, mais également d'enrichir l'ontologie lorsqu'un nouveau tag ne correspond à aucun concept existant. Gruber (2005) a introduit l'idée d'une "ontologie de folksonomie", où chaque événement de tagging a des attributs liés (tag, objet désigné par le tag, tagueur, ressource concernée, etc.).

³Connotea : <http://www.connotea.org/>

⁴<http://scot-project.org/scot/ns>

4 Ontologies et folksonomies sur le Web

Les pratiques du Web 2.0 et les technologies du Web sémantique s'intéressent toutes les deux, mais de façons différentes, à la notion d'un partage social du sens, et ces dernières années ont montré la complémentarité de ces approches (Greaves & Mika, 2008). Les approches de cette section proposent d'exploiter les formalismes du Web sémantique en vue de favoriser l'interconnexion des sites Web à usages communautaires. Ici pas d'ontologie de domaine, mais des annotations reposant sur des ontologies formelles visant à relier efficacement les ressources entre elles. Ces approches peuvent constituer une position intermédiaire qui peut rester indépendante ou non d'un couplage avec des ontologies de domaine.

4.1 Communautés en ligne interconnectées

Le projet SIOC (Breslin *et al.*, 2005) offre un cadre formel et technologique pour décrire les ressources échangées sur les blogs et forums de discussions, et plus généralement sur toute plateforme sociale d'échanges en ligne. L'ontologie proposée utilise d'autres ontologies du Web sémantique, comme SKOS⁵, décrivant les systèmes d'organisation simple de la connaissance, et FOAF⁶ (Brickley & Miller, 2004) pour décrire les identités et les accointances d'un utilisateur. SIOC⁷ décrit les éléments présents dans la plupart des sites communautaires : la notion de "site", comme emplacement de la communauté ; la notion de "post", ou billet pour un blog ; la notion de "forum" ; la notion "d'utilisateur" ; etc. A partir de ce vocabulaire Breslin *et al.* (2005) proposent des extensions pour extraire ces annotations de plateformes courantes (ex : Wordpress⁸).

4.2 Ressources sur le Web

De nombreux autres travaux ont cherché à intégrer plusieurs ontologies dans le but de structurer les données en vue de leur partage.

Hausenblas & Rehatschek (2007) proposent avec "mle" un système permettant le traitement automatique des listes de diffusion par emails, dans le but de constituer une base de connaissances conforme aux standards du Web sémantique. Cette approche utilise SIOC pour représenter les champs des emails (objet, émetteurs, récepteurs, etc.). Ces annotations exprimées en RDF peuvent ensuite être mises à disposition d'un service Web, pour pouvoir être interrogées grâce au langage de requête du Web Sémantique SPARQL⁹.

Le projet Revyu¹⁰ (Heath & Motta, 2007) propose de mettre en pratique les principes du "Web of Data" (Web de Données) pour organiser le partage de revues de contenus culturels (livres, films de cinéma, etc.). Parmi les principes du "Web de Données"

⁵SKOS : Simple Knowledge Organisation System, <http://www.w3.org/2004/02/skos/>

⁶FOAF : Friend Of A Friend, <http://www.foaf-project.org/>

⁷SIOC Ontology <http://www.w3.org/Submission/2007/SUBM-sioc-applications-20070612/>

⁸Logiciel de blog Wordpress, <http://wordpress.org/>

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

¹⁰<http://revyu.com/>

préconisés par Tim-Berners Lee¹¹ nous pouvons citer : l'utilisation d'URIs comme noms et sous forme de lien HTTP pour faciliter l'accessibilité à un maximum d'information sur chaque item indexé et l'interconnexion avec d'autres URIs. Les formalismes du Web Sémantique permettent de réaliser ces principes¹², et Revyu les implémente pleinement : (1) en permettant d'accéder à des données stockées sur d'autres dépôts afin d'éviter la redondance d'informations ; (2) en utilisant RDF pour l'annotation ; et (3) en ne restreignant pas le domaine couvert, puisque Revyu n'impose pas un ensemble défini d'ontologies, mais utilise des sources externes pour catégoriser les items.

D'autres approches visant à structurer sémantiquement des contributions annotées à l'aide de tags ont récemment été proposées, comme GroupMe !¹³ (Abel *et al.*, 2007) ou inter.est¹⁴ (Kim *et al.*, 2007) et consistent à enrichir les services de *bookmarking* social (tel que del.icio.us) à l'aide des technologies du Web sémantique. L'objectif de inter.est est d'intégrer les apports du Web sémantique à un service de *bookmarking* social, et l'ontologie SCOT¹⁵ en est le support. Cette ontologie permet de décrire les termes d'un nuage de tags (ensembles de tags reliés à des ressources et appartenant à des utilisateurs) dans le but d'instrumenter son partage. Le service proposé vise à agréger des nuages de tags, à former des groupes d'échange et à faciliter la recherche de nuages de tags similaires, reflétant ainsi les communautés d'intérêt en germe.

4.3 Wikis sémantiques

Les wikis "sémantiques" ont été parmi les pionniers dans l'utilisation d'ontologies formelles en "soutien" aux folksonomies. SweetWiki (Buffa *et al.*, 2008) en est un exemple : les utilisateurs peuvent éditer et modifier des pages, mais aussi taguer des pages, des images, des vidéos, etc. Les tags sont reliés entre eux au sein d'une folksonomie représentée à l'aide des langages du Web sémantique. Les nouveaux tags sont collectés comme labels (termes) de nouvelles classes qui sont subsumées par défaut par la classe "nouveaux concepts". Ensuite, et selon les principes de contribution et de régulation collective observé au sein du Web 2.0, des utilisateurs peuvent organiser les tags de la folksonomie, les éditer, ajouter des labels dans d'autres langues, créer des relations de synonymie, fusionner des classes, etc.

Les auteurs de pages peuvent utiliser également les tags pour proposer des fonctionnalités de "veille" : chaque utilisateur peut dans sa page personnelle spécifier une liste de tags correspondant aux sujets qui l'intéressent. Par exemple, un utilisateur intéressé par les "wikis", entrera le tag "wiki" dans sa page personnelle : sa page contiendra désormais une liste de liens vers des pages taguées avec "wiki" ou avec une sous-classe de ce mot-clé dans la folksonomie organisée. Cette approche permet de surveiller des ressources qui n'existent pas encore : si aucune page ne parle de "wiki", lorsqu'une page avec ce tag ou un de ses sous-tag est créée, l'utilisateur est notifié. En gardant trace des pages créées et modifiées, des documents tagués et surveillés, et en analy-

¹¹La Recherche (nov 2007) : "L'avenir d'Internet. Tim Berners-Lee : Le Web va changer de dimension"

¹²pour plus de détails voir également : <http://linkeddata.org/>

¹³<http://groupme.org/>

¹⁴<http://int.ere.st/>

¹⁵<http://scot-project.org/scot/ns>

sant le comportement des utilisateurs au cours du temps, il est possible de détecter des réseaux d'accointances ou des communautés d'intérêts. Les possibilités alors offertes sont multiples : trouver la personne la plus active sur un sujet donné, trouver des utilisateurs utilisant le plus de tags équivalents, inférer des relations entre les tags lorsqu'ils partagent des utilisateurs communs, des ressources communes, etc.

4.4 Conclusion partielle

Nous avons ainsi vu l'apport des formalismes du Web sémantique aux services de partages de connaissance fondés sur des contributions de revues (Revyu), ou de *bookmark* (intere.st, GroupMe !) ou encore de pages de wiki (SweetWiki). GroupMe ! et intere.st mettent cependant l'accent sur les échanges sociaux qui accompagnent l'utilisation de ces services, tandis que Revyu se concentre sur l'application des principes du Web de données en favorisant l'interconnexion des ressources. Les wikis sémantiques ont permis quant à eux de montrer l'intérêt des ontologies comme support d'une connaissance "en construction". Buffa *et al.* (2008) proposent ainsi de factoriser la réorganisation des tags-concepts, en permettant à chaque utilisateur de modifier la folksonomie et de fusionner certains tags-concepts, de proposer des synonymes, etc. Ces travaux montrent également les possibilités apportées par le format RDF, en implémentant des solutions de partage des connaissances reposant sur des structures ontologiques.

5 Discussion

Dans cet article nous avons comparé différentes approches visant à rapprocher ontologies et folksonomies pour l'indexation des ressources partagées. Ces approches, souvent complémentaires, peuvent être considérées selon différents axes.

Premièrement, selon un axe structuration/formalisation suivant lequel : (a) soit on cherche à analyser les tags et à inférer des relations sémantiques entre les éléments de folksonomies en vue de leur structuration (section 2) ; (b) soit on cherche à guider l'indexation des interactions et des contributions aux folksonomies grâce à une formalisation à base d'ontologies (sections 3 et 4).

Deuxièmement, on pourra chercher à distinguer les types de ressources annotées et la stratégie adoptée. Abel *et al.* (2007) et Kim *et al.* (2007) cherchent à assister le partage de *bookmarks* avec des ontologies (notamment SKOS, SIOC, et FOAF) visant à assister l'interconnexion des ressources et des personnes au sein des communautés en lignes. Ce même travail a par ailleurs été initié par Breslin *et al.* (2005) dans le cas des blogs et des forums en lignes, et par Hausenblas & Rehatschek (2007) dans le cas des listes de diffusions. Dans le même ordre d'idées, Buffa *et al.* (2008) ont outillé l'édition collaborative de pages de wikis avec des fonctions de tagging social directement formalisées dans les langages du Web Sémantique.

Troisièmement, il est possible de confronter les différents types de groupes sociaux considérés. Les approches de Passant (2007) et de Good *et al.* (2007) s'inscrivent dans des contextes où le champ de connaissances est clairement identifié (un blog d'entreprise ou la biologie). Dans ce cas, il semble plus aisé de concevoir un système centralisé

de contrôle du vocabulaire que dans le cadre des systèmes ouverts de *bookmarking* sociaux (Kim *et al.*, 2007) ou d'agrégation de revues d'objets culturels variés (Heath & Motta, 2007).

Cependant, il est nécessaire de ne pas perdre de vue l'aspect social du partage des connaissances et de s'attacher à concevoir des modèles en adéquation avec les usages. A cet égard Sinha (2006) propose une analyse sociale du tagging. Elle rappelle la distinction introduite par Mathes (2004) entre l'usage privé des services de *bookmarking* sociaux, où le tag a une utilisation personnelle dans le but de retrouver une ressource, et l'usage tourné vers le lien social, où le tag est choisi pour décrire sans ambiguïtés une ressource destinée à être partagée. Sinha montre également que l'annotation d'une ressource à l'aide de plusieurs mots-clés (tags) est cognitivement moins lourde que l'effort du choix d'une catégorie unique. Le tagging se révèle donc plus simple, puisqu'il autorise à taguer avec tous les concepts activés à l'esprit par l'objet à classer.

D'autre part, dans les cas où les vues sur un champ de connaissances ne sont pas consensuelles, Zacklad (2007) suggère le recours à des ontologies semi-structurées, appelées "ontologies sémiotiques" et exprimées selon le modèle "HyperTopic" (Zacklad *et al.*, 2007). Les ontologies sémiotiques nécessitent cependant le recours à un spécialiste des schématisations de la connaissance, et ne constituent donc pas une alternative aussi spontanée d'utilisation que les folksonomies. Néanmoins, elle peuvent être considérées comme une représentation intermédiaire aux ontologies formelles, au sens où l'opérationnalisation des ontologies sémiotiques n'est pas "prolongée par une formalisation référentielle". Il s'agit donc de considérer les processus de négociations, discussions sur les connaissances partagées comme le cœur du problème ; le but n'est pas d'aboutir à un schéma formel opérationnel (ontologie formelle) mais à une cartographie des connaissances ou des réseaux de description (Cahier *et al.*, 2005).

Gruber (2008) différencie intelligence collective et collection d'intelligences. Il donne trois caractéristiques des systèmes de collection d'intelligences : (1) une production de contenu soutenue par les utilisateurs du système, (2) une synergie entre l'homme et la machine en particulier pour couvrir de larges domaines et (3) un gain augmentant avec l'échelle du domaine couvert. Pour passer d'une collection d'intelligences à une intelligence collective, Gruber propose d'ajouter une quatrième caractéristique : l'émergence de connaissances dépassant les connaissances fournies individuellement. Il suggère que cette quatrième caractéristique puisse directement bénéficier de l'intégration des technologies du Web sémantique.

Les approches proposées dans cet article montrent que la nature sociale des échanges de connaissances n'est pas en contradiction avec les possibilités offertes par les systèmes à base d'ontologies formelles. Le point que nous abordons ici est symptomatique de la tension entre l'utilité de représentations formelles dans un système à base de connaissances et l'utilité de représentations informelles dans les interactions quotidiennes entre individus. Parce que les tags sont simples et naturels d'emploi, parce que les folksonomies se constituent d'elles mêmes comme un effet secondaire de l'usage des tags, elles sont considérées, à juste titre, comme extrêmement performantes puisqu'elles utilisent la masse et la diversité des utilisateurs pour organiser la masse et la diversité des ressources d'information. Parce qu'elles ont des méthodologies parfois lourdes, parce qu'elles reposent sur des formalismes complexes, les ontologies sont considérées

comme des objets coûteux et difficilement réconciliables avec des interactions naturelles. Cependant, comme le montrent les exemples abordés dans cet article, il est possible d'utiliser une ontologie pour décrire une folksonomie, l'utilisation des tags ou tout autre activité liée aux usages du Web communautaire. Opposer ontologies et folksonomies n'a donc pas de raison dans l'absolu puisque, pour un scénario donné, elles peuvent avoir des rôles complémentaires à jouer dans une même solution applicative.

Enfin, lorsque l'on regarde les cycles de vie d'une ontologie et d'une folksonomie, les approches présentées ici dévoilent une opportunité de faire cohabiter ces deux objets : (1) les scénarios d'utilisation d'une folksonomie finissent souvent par identifier des cas d'usages où l'on souhaiterait pouvoir exploiter plus avant des structures dans ces ensembles de tags ; (2) la création et la maintenance d'une ontologie doivent toujours trouver des méthodes pour éviter le goulot d'étranglement de l'acquisition des connaissances, par exemple en se nourrissant de folksonomies. Permettre de fusionner certaines structures d'une folksonomie et d'une ontologie, représenter explicitement et maintenir ces points de jonction ; gérer les frictions entre les deux cycles de vie ; assister et automatiser le moins intrusivement possible les échanges entre ces deux objets en arrière plan des usages ; tels sont à notre avis des enjeux intéressants et fertiles pour les systèmes d'information communautaires.

Remerciements au comité de relecture des conférences IC et GECO.

Références

- ABEL F., FRANK M., HENZE N., KRAUSE D., PLAPPERT D. & SIEHNDEL P. (2007). Groupme ! - Where Semantic Web meets Web 2.0. In *ISWC/ASWC*, volume 4825 of *LNCS*, p. 871–878 : Springer.
- ANGELETOU S., SABOU M., SPECIA L. & MOTTA E. (2007). Bridging the Gap Between Folksonomies and the Semantic Web : an Experience Report. *Proc. of ESWC workshop on Bridging the Gap between Semantic Web and Web*.
- BRESLIN J., HARTH A., BOJARS U. & DECKER S. (2005). Towards Semantically-Interlinked Online Communities. In *ESWC 2005*.
- 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., PÉTARD X., LEBOEUF J.-P. & GUITTARD C. (2005). Experimentation of a Socially Constructed “Topic Map” by the OSS Community. *workshop on Knowledge Management and Organizational Memories, IJCAI-05*.
- GOOD B., KAWAS E. & WILKINSON M. (2007). Bridging the gap between Social Tagging and Semantic Annotation : E.D. the Entity Describer. Available from Nature Precedings.
- GREAVES M. & MIKA P. (2008). Semantic Web and Web 2.0. *J. Web Sem.*, **6**(1), 1–3.
- GRUBER T. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, **5**(2), 199–220.

- GRUBER T. (2005). Ontology of Folksonomy : A Mash-up of Apples and Oranges. In *Conference on Metadata and Semantics Research (MTSR)*.
- GRUBER T. (2008). Collective knowledge systems : Where the Social Web meets the Semantic Web. *J. Web Sem.*, **6**(1), 4–13.
- HALPIN H., ROBU V. & SHEPHERD H. (2007). The Complex Dynamics of Collaborative Tagging. In *WWW* : ACM Press.
- HAUSENBLAS M. & REHATSCHEK H. (2007). mle : Enhancing the Exploration of Mailing List Archives Through Making Semantics Explicit. In *Semantic Web Challenge, ISWC*.
- HEATH T. & MOTTA E. (2007). Revyu.com : a Reviewing and Rating Site for the Web of Data. In *ISWC/ASWC*, volume 4825 of *LNCS*, p. 895–902 : Springer.
- JÄSCHKE R., HOTHO A., SCHMITZ C., GANTER B. & STUMME G. (2008). Discovering Shared Conceptualizations in Folksonomies. *J. Web Sem.*, **6**(1), 38–53.
- 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*.
- 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.
- 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*.
- SCHMITZ C., HOTHO A., JÄSCHKE R. & STUMME G. (2006). Mining Association Rules in Folksonomies. p. 261–270.
- SINHA R. (2006). Tagging from Personal to Social : Observations and Design Principles. In *Tagging Workshop, WWW*.
- 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*.
- VANDERWAL T. (2004). Folksonomy Coinage and Definition. <http://www.vanderwal.net/folksonomy.html>.
- ZACKLAD M. (2007). Classification, Thésaurus, Ontologies, Folksonomies : Comparaisons du Point de vue de la Recherche Ouverte d'Information (ROI). In *CAIS/ACSI*.
- 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.