



HAL
open science

Folksonomies, Thésaurus et Ontologies : trois artefacts combinés dans la structuration des données du Web

Fabien Gandon, Alexandre Passant

► To cite this version:

Fabien Gandon, Alexandre Passant. Folksonomies, Thésaurus et Ontologies : trois artefacts combinés dans la structuration des données du Web. Classer les archives et les bibliothèques, Presses universitaires de Rennes, 2015, 10.4000/books.pur.88593 . hal-02920553

HAL Id: hal-02920553

<https://inria.hal.science/hal-02920553>

Submitted on 24 Aug 2020

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.

Folksonomies, Thésaurus et Ontologies : trois artefacts combinés dans la structuration des données du Web

Fabien L. Gandon, (INRIA) Institut national de Recherche en Informatique et en Automatique, Sophia Antipolis

Alexandre Passant, (DERI) Digital Enterprise Research Institute, National University of Ireland, Galway

Abstract

Au cœur de nos systèmes documentaires, de nos bibliothèques numériques, de nos systèmes d'information, du Web 2.0 et du Web Sémantique, les ontologies, les thésaurus et les folksonomies sont trois des structures de données qui participent à l'indexation des contenus. Parfois confondus, parfois opposés, nous montrons que ces trois « artefacts cognitifs » qui se répandent actuellement dans les applications Web répondent à des besoins différents et peuvent très bien être combinés au sein d'une même application pour permettre différentes fonctionnalités, offrant ainsi différents modèles et permettant différentes pratiques pour l'indexation de contenus en ligne.

Introduction

L'avènement des systèmes de gestion de contenu sur le web, qu'il s'agisse de systèmes d'information d'entreprise, de bibliothèques numériques ou de sites sociaux du Web 2.0, permet la création de plus en plus simples de contenus textuels ou multimédia en ligne. Or, si la publication est rendue plus simple, il convient d'offrir des méthodes efficaces d'indexation pour retrouver les différents contenus produits et faire face à la surcharge d'information causée par certains outils, notamment dans le contexte du Web 2.0 et des applications sociales en ligne (blogs, gestion de favoris, etc.).

Pour ce faire, les ontologies, les thésaurus et les folksonomies sont trois des structures de données qui participent à cette indexation des contenus. Parfois confondus, parfois opposés, nous montrons ici que ces trois « artefacts cognitifs » qui se répandent actuellement dans les applications Web répondent à des besoins différents et peuvent très bien être combinés au sein d'une même application pour permettre différentes fonctionnalités, offrant ainsi différents niveaux pour l'indexation de contenus en ligne. Ainsi, nous argumentons que non seulement ces trois artefacts ne sont pas opposés, mais qu'au contraire ils peuvent être utilisés en combinaison, et peuvent bénéficier chacun des apports de l'autre.

Tout d'abord, nous présenterons les définitions de ces trois artefacts cognitifs, appuyées par différents exemples, et discuterons les étapes nécessaires pour passer d'un artefact à l'autre. Dans un second temps, nous présenterons différentes raisons qui conduisent trop souvent à les considérer comme opposés. Enfin, dans une troisième partie, nous présenterons plusieurs initiatives qui permettent de faire le lien entre ces artefacts, et notamment qui permettent : (1) l'extraction d'un type d'artefact à partir d'un autre type, notamment la construction d'ontologies et de thésaurus à partir de folksonomies et (2) l'utilisation d'ontologies pour représenter folksonomies et thésaurus, permettant de bénéficier d'un type de représentation particulier pour en représenter un autre.

Définitions

Folksonomies

Si les systèmes documentaires tels que les bibliothèques numériques font généralement appel à des systèmes structurés d'indexation, gérés par des experts du domaine (tels que les thésaurus que nous présenterons par la suite), le Web 2.0¹ a introduit une pratique basée sur la catégorisation des contenus par les utilisateurs eux-mêmes. Celle-ci repose en générale sur l'association aux contenus en ligne de mots-clés libres (aussi bien en type, nombre ou langue), encore appelés tags. Si le tag est couramment un terme, rien ne s'oppose à l'utilisation de tout type de signe. De par son rattachement à un contenu existant (billet de blog, photo, vidéo, etc.), un tag peut essentiellement être vu comme une métadonnée supplémentaire associée à une ressource. Cependant, alors qu'un outil de blog associe automatiquement à un billet la date de création de celui-ci et le nom de son auteur, qu'une photo possède ses métadonnées EXIF pour identifier ses caractéristiques, les métadonnées générées ici sont contrôlées et personnalisées par la foule des utilisateurs eux-mêmes, et l'on parle ainsi de métadonnées sociales.

De plus, à l'utilisation de ces tags est liée la pratique d'étiquetage ou de tagging: association par un utilisateur d'un tag à une ressource donnée (billet de blog, photo ...). Cette action² forme ainsi une relation tripartite³ qui peut se représenter par le prédicat ternaire : *Tagging(Utilisateur, Ressource, Tag)*, où:

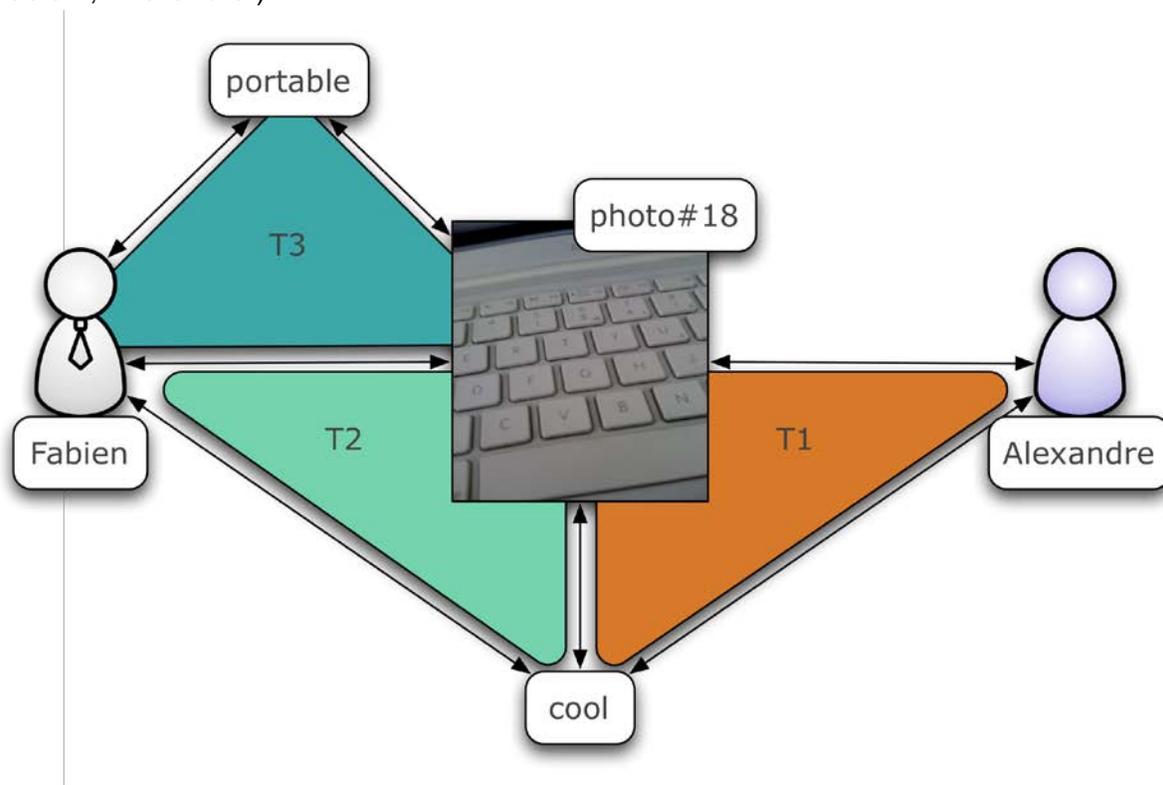
- *Utilisateur* correspond à la personne qui effectue l'action de taguer (ex. "Fabien", "Alexandre", etc.) ;
- *Ressource* correspond à la ressource annotée (ex. un billet de blog, une photo, un clip, etc.) ;
- *Tag* correspond au signe utilisé pour taguer, en général un terme (ex. "portable", "cool", "mer",":-)");
- *Tagging* correspond à l'action liant ces trois éléments (ex. "alexandre" tague "photo#18" avec le terme "mer").

¹ O'Reilly T., *What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software*, 30 September 2005. <http://www.oreillyn.com/lpt/a/6228>

² Halpin H., Robu V., Shepard H., *The Dynamics and Semantics of Collaborative Tagging*. Actes Semantic Authoring and Annotation Work- shop (SAAW'06), volume 209 of CEUR Workshop Proceedings. CEUR-WS.org, 2006.

³ Mika P., *Ontologies Are Us: A Unified Model of Social Networks and Semantics*. Actes de International Semantic Web Conference (ISWC 2005), volume 3729 of Lecture Notes in Computer Science, pages 522–536. Springer, 2005

La figure qui suit illustre ainsi trois actions de *Tagging* (T1, T2, T3) associée à une même Ressource (“photo#18”) faisant intervenir deux Tags (“portable”, “cool”) et deux Utilisateurs (“Fabien”, “Alexandre”).



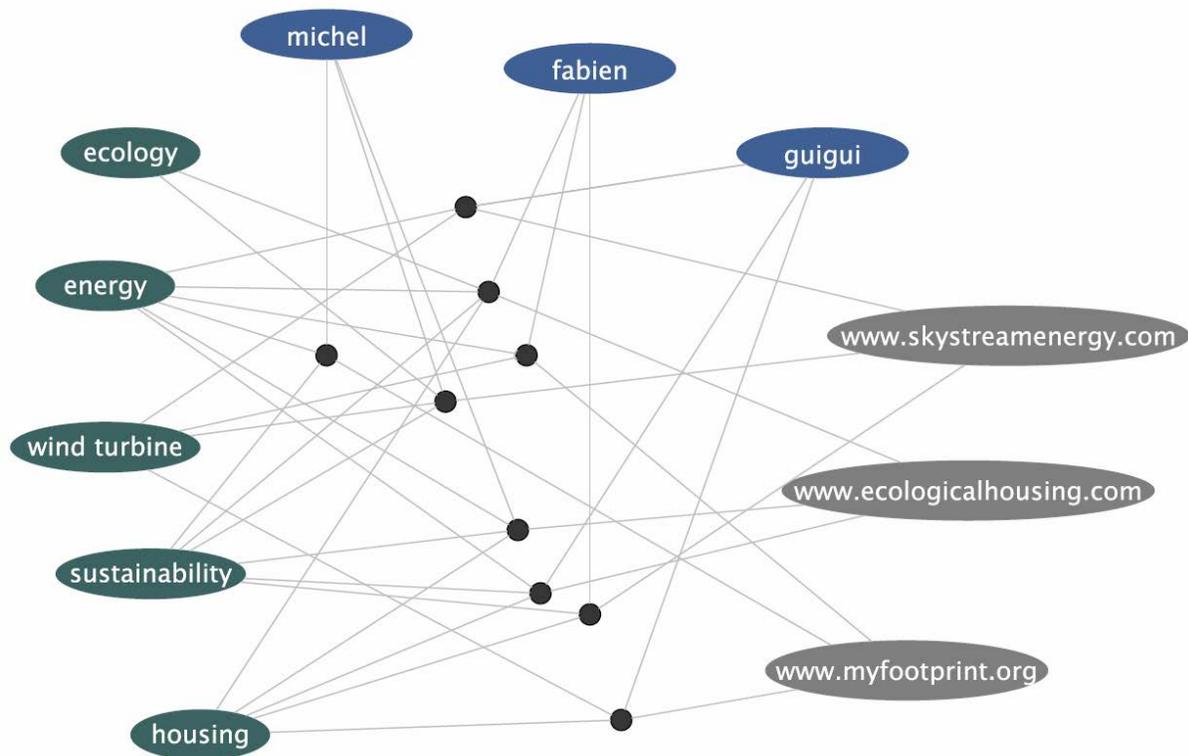
Le graphe tripartite joignant toutes les actions de tagging au sein d’un espace donné (site Web, plate-forme de blogs ...) forme ce que l’on appelle une folksonomie⁴, terme issu du croisement entre *folks* (les gens) et *taxonomy* (taxonomie) et dont la pratique a fait l’objet de nombreuses publications ces dernières années⁵.

La figure suivante donne un exemple du graphe tripartite d’une folksonomie dans le contexte d’une étude du tagging à l’Ademe⁶ avec neuf actions de tagging, par trois utilisateurs à propos de trois ressources et utilisant cinq tags.

⁴ Vander Wal T., *Folksonomy Coinage and Definition*, 2007. <http://www.vanderwal.net/folksonomy.html>

⁵ Golder S., Huberman B., *Usage patterns of collaborative tagging systems*, *Journal of Information Science*, 32(2):198–208, 2006 ; Mika P., *loc. cit.* ; Halpin H. *loc. cit.*

⁶ Limpens F., *Multi-points of view semantic enrichment of folksonomies*, thèse de doctorat, Inria, Université Nice - Sophia Antipolis, 25 Octobre, 2010



La simplicité de cette approche de classification de contenus fait la force des systèmes à base de tags: pas de vocabulaire prédéfini à appréhender par l'utilisateur, évolution constante par l'ajout de nouveaux termes, etc. Cependant, ceux-ci souffrent de nombreux défauts en termes de traitement automatisé et notamment pour la recherche d'information, causés aussi bien par les problèmes d'ambiguïté ou de synonymie des mots-clés que par leur nature totalement plate et l'absence de liens entre tags^{7,8}. D'une part, un même tag peut faire référence à deux idées distinctes ("apple", "glace", "jaguar" etc.). D'autre part, différents tags peuvent faire référence à la même idée ("voiture", "automobile", "caisse"). Enfin, rien ne permet non plus d'établir un lien direct entre deux tags ("véhicule", "voiture", "coupé", "407").

A contrario, une des forces des folksonomies se situe dans leur utilisation en termes de navigation et dans les possibilités qu'elles offrent pour la découverte de nouvelles informations. L'organisation des liens entre ressources, tags et utilisateurs forme en effet un graphe dans lequel il est possible de naviguer renforçant ainsi la sérendipité, qu'il s'agisse de découvrir de nouveaux documents, de nouveaux thèmes ou de nouveaux utilisateurs, du fait du modèle tripartite exposé auparavant. La popularité des tags dans une folksonomie est d'autre part rendue visible par l'utilisation de nuages de tags ou *tagclouds*, offrant également un autre mode de navigation permettant d'avoir un aperçu du champ lexical associé à une folksonomie et des tendances émergentes au sein d'un espace informationnel donné, si l'on ne considère par exemple que les dernières actions de tagging. La figure qui suit présente ainsi la page d'accueil

⁷ Mathes A., *Folksonomies: Cooperative Classification and Communication Through Shared Metadata*, December 2004

⁸ Passant A., Sta J-D, Laublet P., *Folksonomies, Ontologies and Corporate Blogging*. Actes Blogtalk Conference (Blogtalk Reloaded). Books on demand, 2006

du gestionnaire collaboratif de bibliographies BibSonomy⁹, ou l'on peut observer le nuage des tags les plus récents à droite dans l'encadré.

bookmarks RSS BibTeX XML

Live demos of Zimbra's open source email and group calendar software
to GroupWare by macek on Sep 2, 2010, 4:55 PM
spam

Online Python Tutor: Write Python code online and single-step through its execution
This application supports the core Python 2.5 language, with no module imports or file I/O. It's meant to be used as a platform for creating programming tutorials.
to programming python by toni on Sep 2, 2010, 4:06 PM
spam

KoopTech » Titelgeschichte » Buzz Kill: Und niemand merkt, dass du verschwunden bist
to koop-tech social-networking by bunke37 on Sep 2, 2010, 3:32 PM
spam

J Web Semantics VS the Semantic Web? — The problem of keyness
to paper by voj on Sep 2, 2010, 3:21 PM
spam

www.um.org
to community modeling user by kasimiro on Sep 2, 2010, 3:20 PM
spam

publications RSS BibTeX RDF more

A Genetic Algorithm for Function Optimization: A Matlab Implementation
Christopher Houck North and Jeffery A. Joines and Michael G. Kay and Christopher R. Houck and Christopher R. Houck (1996)
to 1996 Matlab algorithm evolutionary optimization by thorade on Sep 2, 2010, 4:24 PM
URL | BibTeX | spam

A Proposal for Publishing Data Streams as Linked Data - A Position Paper
Davide Barbieri and Emanuele Della Valle Proceedings of the Linked Data on the Web LDOW2010 Workshop, co-located with WWW2010, (2010)
to Data Linked Open RDF Streams by dadeit on Sep 2, 2010, 4:14 PM
URL | BibTeX | spam

The Experience of Realizing a Semantic Web Urban Computing Application
Emanuele Della Valle and Irene Celino and Daniele Dell'Aglio T. GIS14(2):163-181(2010)
to computing semantic urban web by dadeit and 1 other user on Sep 2, 2010, 4:12 PM
BibTeX | spam

Towards Treating GIS as Virtual RDF Graphs
Emanuele Della Valle and Hafiz Muhammad Qasim and Irene Celino Proceedings of 1st International Workshop on Pervasive Web Mapping, Geoprocessing and Services WebMGS 2010, (2010)
to pervasive web mapping by dadeit on Sep 2, 2010, 4:11 PM
URL | BibTeX | spam

• **Feature of the week:**
Configurable displayed resources on Aug 9, 2010

☐ **busy tags**
(alpha | freq) (cloud | list)

2009 2010 algorithm analysis
android blog book
classification clustering
conference css data database
dblp design
development editor
education elearning evaluation
facebook film folksonomy free
generator google high history
howto html identity
information innovation
interface internet iphone java
javascript journal knowledge
language latex learning lib
library links linux lis literature
management

Thésaurus

Si les folksonomies sont des artefacts totalement non-structurés, les thésaurus suivent une approche différente. L'objectif d'un thésaurus est de constituer un vocabulaire normalisé et d'organiser la liste de termes (sans forcément les définir) dans le but notamment d'indexer un corpus de documents et de faciliter les recherches dans ce corpus.

Le thésaurus permet à la fois de guider l'indexation en suggérant des termes à utiliser et d'enrichir des requêtes ou de proposer des reformulations. Pour cela on différencie deux types de termes reliés au sein du thésaurus par trois types de relations :

- Les termes descripteurs sont reliés aux termes synonymes non descripteurs par des relations d'équivalences ("voiture == "automobile").
- Les termes descripteurs sont organisés en une hiérarchie par une relation générique-spécifique ("véhicule" > "voiture").
- Les termes descripteurs sont reliés en champs thématiques, sujets connexes et termes associés par une relation d'association ("voiture" → "conducteur").

L'indexation s'impose d'utiliser les termes descripteurs les plus spécifiques pour ensuite utiliser la hiérarchie et les équivalences du thésaurus afin d'augmenter la précision et le rappel des recherches. Par exemple les documents seront indexés par "206" et identifiables lors d'une recherche sur "véhicule", "voiture" ou "automobile". Les ontologies informatiques (définies après) et les thésaurus sont souvent des notions qui se mélangent à tort ou à raison: domaines proches, structures hiérarchiques ressemblantes, applications proches, etc. Le thésaurus a

⁹ <http://bibsonomy.org>

pourtant une différence forte avec les ontologies : le thésaurus est un recueil de termes d'un domaine et il organise directement ces termes.

Ontologies

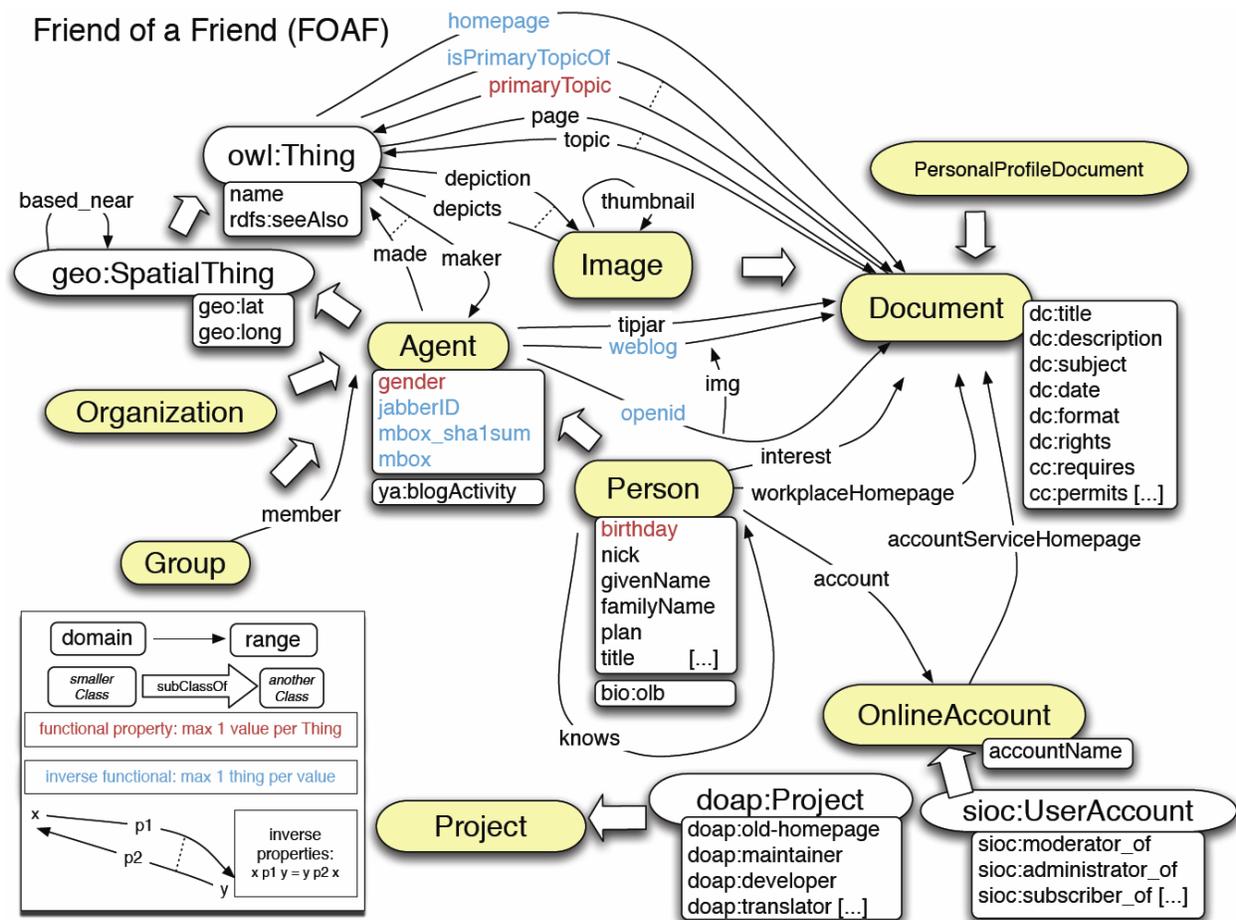
Le terme 'ontologie' fut officiellement emprunté par l'informatique¹⁰ à la philosophie au début des années 1990. En philosophie, l'Ontologie est une branche qui étudie l'être en tant qu'être, l'existence en général. En informatique le terme d'ontologie¹¹ est entré par le domaine de l'intelligence artificielle symbolique qui s'intéresse à la représentation des connaissances. Une ontologie informatique est un objet logiciel, une représentation de propriétés générales de ce qui existe pour une application dans un formalisme supportant un traitement rationnel. C'est le résultat d'une formulation exhaustive et rigoureuse de la conceptualisation d'un domaine, comme le rappelle sa définition: "An ontology is a specification of a conceptualization."¹² Cette conceptualisation formulée est nécessairement partielle et son degré de formalisation varie avec l'usage qui en est envisagé, et le domaine abordé dans la conceptualisation, qui peut être de nature plutôt abstrait (ontologie du temps) ou au contraire très précis (ontologie de l'automobile). C'est parce qu'elles ont cet objectif de description de l'existant et de ses catégories que les ontologies informatiques ont emprunté leur nom à l'Ontologie philosophique. Une ontologie informatique fournit un référentiel pour construire un système à base de connaissances et concevoir des inférences (recherche d'information, raisonnement) et assurer l'interopérabilité entre les systèmes la partageant. Par exemple, la figure qui suit (par Dan Brickley) représente l'ontologie FOAF (Friend Of A Friend¹³) permettant la représentation des personnes et de leurs réseaux sociaux.

¹⁰ Gruber T. R., *Towards Principles for the Design of Ontologies Used for Knowledge Sharing*. International Journal Human-Computer Studies, 43(5-6):907-928, 1995

¹¹ Gandon F., *Ontologies in Computer Science*, in *Ontology Theory, Management and Design: Advanced Tools and Models*, Ed. Faiez Gargouri, Wassim Jaziri, Pages 1-26, ISBN10: 1615208593

¹² Gruber T.R., *A Translation Approach to Portable Ontology Specifications*. Knowledge Acquisition, 5(2), 199-220, 1993

¹³ Brickley D., Miller L., *FOAF Vocabulary Specification 0.97*. Namespace Document 1 January 2010 - 3D Edition 1 January 2010, 2010 <http://foaf-project.org>



Une ontologie inclut couramment une organisation hiérarchique des concepts pertinents et des relations qui existent entre ces concepts, ainsi que des règles et axiomes qui les contraignent. La taxinomie ou hiérarchie de types est très souvent le squelette minimal de l'ontologie. On appelle "subsomption" le fait de placer une catégorie sous une autre ; c'est aussi le lien qui en résulte entre la sous-catégorie subsumée et la catégorie mère. Par exemple la catégorie "véhicule" peut subsumer la catégorie "voiture". L'importance de l'organisation taxinomique se justifie par le fait que la classification ou identification (le fait de déterminer si quelque chose appartient à une classe) et la catégorisation (le fait d'identifier les catégories existantes) sont des inférences élémentaires que nous faisons à longueur de journée. Ce squelette taxinomique est aussi la principale source de confusion entre les ontologies et d'autres structures ayant une hiérarchie pour squelette, tels les thésaurus présentés auparavant. Il est aussi très souvent la source de confusion entre taxinomie et ontologie alors que si les connaissances taxinomiques sont effectivement des connaissances ontologiques, les connaissances ontologiques ne se réduisent pas aux connaissances taxinomiques. Ainsi, au delà de la subsomption, d'autres propriétés sont utilisées dans les ontologies pour représenter ces conceptualisations. Par exemple, une propriété "a_pour_composant" liera la catégorie voiture à la catégorie moteur. Une partonomie, hiérarchie méréologique basée sur la composition et non la subsomption (ex. toute voiture a un moteur), est une connaissance ontologique mais pas taxinomique. Également, lors d'utilisation d'ontologie, il convient de distinguer les classes (concepts) et les

instances (réalisation de ces concepts). Par exemple, si “voiture” est considérée comme une classe, on aura tendance à définir “la voiture immatriculée 94BRN06” comme instance de cette classe.

Le cycle de vie d'une ontologie informatique pose de nombreux problèmes comme tout objet logiciel: détection des besoins, conception, gestion et planification, évolution, diffusion, utilisation, évaluation. Les méthodologies d'ingénierie qui se sont développées autour de cet objet font qu'il est souvent perçu comme coûteux et contraignant alors qu'il existe de nombreuses petites ontologies utiles, virales et peu coûteuses, comme celles que nous présenterons plus loin.

De la folksonomie au thésaurus, du thésaurus à l'ontologie

Comme le montrent les définitions précédentes, folksonomies, thésaurus et ontologies sont trois modes de classification des données, offrant chacun des niveaux d'expressivité différents. Si l'on considère ainsi le simple exemple d'une folksonomie contenant les tags “véhicule, avion, aéroport, réacteur, jet, air force one”, on peut s'intéresser à construire un thésaurus tel que “véhicule” > “avion” > “air force one”, “avion” > “réacteur”, “avion” → “aéroport” et “avion” == “jet”. On retrouve bien des termes descripteurs et synonymes, une relation générique-spécifique et un champ thématique. Enfin, si l'on souhaite structurer à nouveau cette représentation plus loin qu'une simple relation générique-spécifique, il est possible de définir une ontologie contenant les classes organisées “véhicule” > “avion” > “jet” (relation générique-spécifique) et “avion” “a_pour_composant” “réacteur”, “air force one” étant ici considéré comme une instance de la classe “avion”

Opposition ou complémentarité ?

A la lecture des définitions précédentes, on note que chaque notion s'intersecte avec les autres tout en ayant des différences fortes. Si nous avons discuté brièvement le passage de l'une à l'autre d'un point de vue conceptuel, les intersections entre ces notions et le manque de précision dans leurs définitions ont parfois conduit à des mises en opposition radicales (notamment dans le cas des ontologies et des folksonomies) ou au contraire à les amalgamer un peu trop rapidement (notamment dans le cas des ontologies et des thésaurus). Nous avons très tôt travaillé à éviter ces oppositions¹⁴ et nous allons ici revisiter les principales mises en opposition entre ces notions et tempérer ces positions.

Faux-semblant

Nous avons donc trois « artefacts cognitifs » qui se constituent actuellement dans les applications Web: la folksonomie, le thésaurus et l'ontologie. Les pratiques du Web 2.0 ont surpris par leur efficacité. Le *social tagging*, notamment, utilisé avec succès dans des

¹⁴ Gandon F., *Le Web sémantique n'est pas antisocial*, Ingénierie des Connaissances, Semaine de la Connaissance, Nantes, pp. 131-140, 20-30 Juin 2006

applications comme Flickr¹⁵, Delicious¹⁶ ou Youtube¹⁷ produit des folksonomies dans des proportions et avec une rapidité que l'on aimerait pouvoir transposer à l'acquisition des connaissances.

Les contributions d'un utilisateur à une folksonomie restent légères et découlent de son usage : les tags employés par les utilisateurs sont collectés et analysés en tâche de fond, et la folksonomie évolue donc en fonction de ces actions utilisateur. Les contributions aux ontologies sont très souvent directes et demandent des actions dédiées : les concepts et axiomes doivent être validés, organisés, définis etc. Les contributions aux thésaurus suivent souvent un processus strict permettant de valider, normaliser et insérer les nouveaux termes dans la structure, supervisé par des experts du domaine.

On peut dès lors être tenté de dire que les folksonomies offrent une alternative efficace et légère aux deux autres structures. Cependant, lorsque Flickr a dépassé les 20 millions de tags uniques début 2008, la question de gérer cette collection de tags et de l'exploiter efficacement est devenue centrale à l'application.

Dans une ontologie informatique, pour exprimer, communiquer un concept, nous choisissons une représentation symbolique, souvent linguistique et verbale, parfois iconique. Par exemple il existe un concept dont la définition est « qui sert de lien entre plusieurs entités » et des représentations linguistiques de ce concept sont les termes « médiateur », « médiatrice » ou « intermédiaire ». Nous dissociions donc les concepts et leurs libellés. Un terme n'est pas un concept, et vice-versa. Un terme peut être ambigu, alors qu'un concept n'a qu'un seul sens, une seule définition. Les ontologies ne sont pas centrées sur les termes mais sur les notions. Ainsi, pour prendre l'exemple d'un formalisme du web sémantique, une classe RDFS¹⁸ n'est pas un mot mais un type auquel on peut attacher des labels qui, eux, peuvent être considérés comme des termes. Un schéma RDFS ne formalise pas des termes mais normalise l'identification unique de notions documentées. Plusieurs labels pouvant être associés, on peut donc considérer que plusieurs chemins sémiotiques (en l'occurrence, symboliques) sont documentés pour arriver à cette notion. Il faut alors gérer les problèmes de synonymie (un concept dénoté par plusieurs termes) et d'homonymie (un terme dénotant plusieurs concepts) que nous avons évoqué en présentant les folksonomies.

Or le tag n'est qu'un signe, le plus souvent un terme, utilisé pour marquer une ressource et cela n'en fait pas un concept, un descripteur normalisé ou une classe: ce que l'on gagne en facilité de captation on le perd en facilité d'exploitation et en capacité d'automatisation. Comme le soulève également¹⁹: "une folksonomie représente simultanément une partie du pire et du meilleur dans l'organisation de l'information".

Le terme libellant un concept dans une ontologie et le terme utilisé comme tag dans une folksonomie sont un premier point de rencontre entre ces deux objets de l'informatique. Dans un cycle de construction d'ontologie, le tag se rapproche plus d'un nouveau type de candidat

¹⁵ <http://flickr.com>

¹⁶ <http://delicious.com>

¹⁷ <http://youtube.com>

¹⁸ RDF Vocabulary Description Language 1.0: RDF Schema. W3C Recommendation 10 February 2004, World Wide Web Consortium, 2004. <http://www.w3.org/TR/rdf-schema/>

¹⁹ Mathes A., *loc. cit.*

terme²⁰ dont il nous faut exploiter les spécificités si l'on veut retrouver une structure terminologique ou une conceptualisation. Nous ne disons pas ici qu'il faut systématiquement chercher à faire évoluer le tag vers un label de concept ontologique mais que la mise en correspondance de "termes tags" avec des "termes labels de concepts, de classes, de propriétés" est un premier point de rapprochement identifiable entre ces deux approches.

Nonobstant ce pont, les ontologies, thésaurus et folksonomies ne répondent pas aux mêmes besoins, ne permettent pas les mêmes traitements et ont chacun une place différente dans l'architecture d'une application. Les ontologies sont utilisées avec succès pour permettre l'interopérabilité de systèmes d'information. Une folksonomie ne peut pas remplir simplement ce rôle et un thésaurus ne remplira qu'un type précis de scénarios par exemple l'alignement d'index. A l'inverse, les folksonomies sont utilisées avec succès pour permettre l'organisation collective de ressources documentaires à grande échelle, de façon libre et flexible répondant rapidement à des applications ne nécessitant pas forcément de formalisation ontologique ou pour lesquelles une formalisation serait trop coûteuse, complexe ou longue. L'acte de taguer est lui-même moins contraignant que l'acte de catégoriser puisque dans le premier cas l'utilisateur est libre de choisir les termes qu'il veut et dans le deuxième il doit identifier dans une hiérarchie qu'elle catégorie existante ou à suggérer s'applique.

A la marge, la fragilité de ce pont sémiotique entre les trois structures est pour nous symptomatique de l'atrophie chronique du niveau sémiotique au dessus des formalismes de représentation à base d'ontologies ainsi que d'une articulation logique entre ces deux niveaux. Nous avons par ailleurs discuté²¹ le besoin d'une logique sémiotique et pragmatique pour gérer les interactions entre des représentations de connaissance *ex machina* et des interprétations en entrée comme en sortie d'interactions *in vivo* avec des utilisateurs.

Faux combat

Les ontologies se définissent par le type de leur contenu. Les folksonomies se définissent par leur moyen d'obtention. Le thésaurus se définissent par les deux. Il n'y a donc aucune incompatibilité théorique entre ontologie et folksonomie. Une question intéressante au contraire, est de voir ce que chaque objet peut apporter aux autres.

Par exemple, peut-on obtenir des ontologies par le biais des folksonomies notamment pour des ontologies de domaines, et obtenir ainsi des folks-ontologies²²? Dans cette nouvelle optique, l'ontologie n'est pas la responsabilité unique d'un ontologue mais celui-ci peut participer à l'animation d'une communauté qui se fédère autour de l'utilisation des applications de cette ontologie. La communauté, par son activité, nourrit le cycle de vie de cette ontologie, qui devient de facto un artéfact social. Dans le meilleur des cas, l'évolution de l'ontologie devient un effet secondaire de l'activité normale de la communauté. Du point de vue de l'ingénierie ontologique, il s'agit non pas de faire rentrer l'utilisateur dans une boucle de revue

²⁰ Aussenac-Gilles N., Biebow B. & Szulman S., Revisiting ontology design: A methodology based on corpus analysis. Actes European Workshop on Knowledge Acquisition, Modeling and Management, p. 172–188, London, UK: Springer-Verlag, 2000

²¹ Gandon F., *Generating Surrogates to Make the Semantic Web Intelligible to End-Users*, IEEE/WIC/ACM International Conference on Web Intelligence, Compiegne University of Technology, September 19-22, 200

²² Van Damme C., Hepp M., Siorpaes K., *FolksOntology: An Integrated Approach for Turning Folksonomies into Ontologies*. Actes ESWC'2007 workshop Bridging the Gap between Semantic Web and Web 2.0, 2007

ou d'orchestration mais d'assigner des tâches à une masse d'utilisateurs de façon à traiter sur une grande échelle des problèmes notoirement difficiles comme la détection d'un concept dans une ressource multimédia, l'organisation des concepts ou la désambiguïsation. A titre d'exemple pionnier, nous citerons le jeu OntoGame²³. Faire intervenir une communauté dans la conception d'une ontologie est déjà une approche connue. Outiller la communauté pour que par son activité normale elle fasse naturellement vivre les ontologies qui la concernent reste une perspective. Les cycles de vie des folksonomies et des ontologies sont différents et la question de la gestion de leur symbiose nous paraît une perspective prometteuse: comment exploiter les folksonomies pour faire émerger des ontologies (analyse statistique des réseaux de tags, utilisation de ressources linguistiques ou d'ontologies existantes et techniques d'alignement, analyse linguistique, analyse des usages)? Dans cette perspective la folksonomie peut se situer comme une représentation intermédiaire (au sens de METHONTOLOGY²⁴) dans l'obtention d'ontologies et de thésaurus.

Une autre perspective est celle qui consiste à remarquer que les folksonomies et les thésaurus nécessitent des langages pour les représenter, les traiter, les échanger. Les ontologies peuvent alors être utilisées comme des méta-modèles fournissant les primitives de représentations de folksonomies et de thésaurus. Un cas particulier dans la mise en opposition de ces objets est le contexte du Web sémantique et du Web social, le premier fournissant des langages de représentation comme RDF/S, OWL et SKOS, le deuxième des applications de social tagging générant des folksonomies. Le Web demande, entre autres, la prise en compte d'une *dimension sémiotique* (en tant que système signifiant mobilisant des signes dans toutes leurs formes pour des représentations à consommation humaine), d'une *dimension pragmatique* (en tant que système sémiotique immergé dans des usages multiples qui en influencent l'interprétation) et d'une *dimension sociale* (en tant qu'espace d'interaction) et rien dans le Web sémantique ne s'y oppose, bien au contraire: l'utilisation d'ontologies formelles permet, là où ceci est possible et intéressant, de figer la sémantique de primitives choisies, par exemple pour assurer un canal d'interopérabilité. Ainsi, SKOS (Simple Knowledge Organisation System²⁵) est un excellent exemple: ce langage est utilisé pour représenter et partager des classifications, des glossaires, des thésaurus, des *folksonomies*, etc. et il est entièrement écrit dans les formalismes du Web Sémantique, i.e. orientés ontologie.

Faux procès

Nous reprenons enfin ici quelques idées reçues rencontrées dans la littérature et que nous souhaitons réfuter.

Le piège du domaine.

Cette idée reçue consiste à croire qu'une application à base d'ontologies en biologie nécessite de modéliser le domaine de la biologie dans une ontologie. Il s'agit de l'idée fausse que le

²³ Siorpaes K., Hepp M., *OntoGame: Weaving the Semantic Web by Online Gaming*, European Semantic Web Conference (ESWC) 2008, Springer LNCS, Tenerife, Spain, June 2008 et <http://ontogame.sti2.at>

²⁴ Fernández-López M, Gómez-Pérez A, Juristo N., *METHONTOLOGY: From Ontological Art Towards Ontological Engineering*. Spring Symposium on Ontological Engineering of AAAI. Stanford University, California, 1997, pp 33–40

²⁵ <http://www.w3.org/2004/02/skos/>

domaine d'application est aussi le domaine de modélisation. Il nous semble donc important de discuter la notion de domaine formalisation: le domaine à formaliser dans l'ontologie n'est pas forcément le domaine d'application; la formalisation est essentiellement spécifiée par la tâche. Ainsi si la tâche est l'assistance à l'alignement de plusieurs terminologies médicales, la formalisation se focalisera peut-être sur des primitives linguistiques (ex: terme, synonyme de, hyponyme de, acronyme de, etc.) permettant de représenter et comparer les différentes terminologies médicales. En d'autres termes, on ne retrouve pas forcément les notions du domaine d'application dans l'ontologie.

S'il est vrai que l'ontologie va chercher à stabiliser la définition de certaines notions notamment pour en faire des primitives de représentation et assurer leur traitement et l'interopérabilité, cette contrainte ne contamine pas pour autant toutes les notions du scénario applicatif considéré. Ainsi, l'interopérabilité recherchée par exemple dans une tâche ou un scénario n'est pas forcément au même niveau que l'interopérabilité recherchée dans une autre tâche ou un autre scénario au sein d'une même application ou entre deux applications. De plus dans un même scénario l'interopérabilité (computationnelle) peut être à un certain niveau ou sur un certain domaine (ex: actes du langage) et l'intercompréhension (ou interopérabilité entre utilisateurs) peut porter sur un autre domaine (ex: arguments de conception échangés et organisés par ces actes du langage).

La question du domaine de modélisation n'est répondue que par l'analyse des scénarios d'usage. En particulier, une ontologie peut s'intéresser à fournir les primitives permettant de capturer une folksonomie, un lexique, un thésaurus, etc. pour une application dans un domaine arbitraire ; le domaine de modélisation est alors celui des folksonomies, lexiques, thésaurus, etc.

La hantise de l'ontologie universelle

L'un des arguments les plus utilisés contre les approches à base d'ontologies (et par déclinaison contre le Web Sémantique dans son ensemble) dit qu'il est impossible de construire une ontologie universelle et donc que les ontologies sont vouées à l'échec.

Confondre la famille de modèles que représentent les ontologies avec certaines initiatives par ailleurs très instructives qui se sont intéressées à construire d'immenses ontologies généralistes est d'autant plus dommageable que, par définition, une ontologie part de l'explicitation d'une conceptualisation et donc d'un compte-rendu partiel de règles que nous reconnaissons régir notre perception de la réalité.

Par définitions les ontologies matérialisent un point de vue forcément limité et généralement volontairement focalisé. De plus, de nombreuses approches et formalismes intègrent cet état de fait en proposant des mécanismes d'extension, d'alignement et de mise en correspondance entre plusieurs ontologies. Ainsi, RDF, le modèle à la base de tout le Web Sémantique, est par nature un système ouvert et extensible où tout le monde peut dire tout ce qu'il veut à propos de toute chose. On peut tout dire et son contraire en même temps. RDFS permet à une ontologie légère de réutiliser et d'étendre tout ou partie d'autres ontologies. OWL permet de dire que telles ou telles primitives ontologique sont équivalentes à telles ou telles autres primitives d'un autre schéma. Ces approches à base d'ontologies sont conçues sur l'idée que la représentation des données est ouverte à de multiples points de vue et que de multiples ontologies seront proposées pour les structurer, parfois complémentaires, parfois redondantes, parfois

différentes, parfois incompatibles. Cet aspect est même un domaine de recherche et développement à part entière comme le montre l'initiative OAEI²⁶.

Opposer folksonomies et ontologies et dans la foulée web social et web sémantique est, au mieux, contre-productif. Le Web Sémantique ne cherche en aucun cas l'ontologie unique, comme le rappelle, par exemple, la discussion par blogs interposés entre Jim Hendler et Clay Shirky²⁷. Au contraire, il intègre par essence le fait que la diversité des métadonnées est notre meilleure arme pour contrôler la diversité des ressources d'information. Ainsi opposer ontologies et folksonomies sur leur capacité d'ouverture, de flexibilité d'extensibilité revient à se couper des voies ouvertes par des systèmes hybrides. Multiplier les systèmes de métadonnées et favoriser leur ouverture et leur interopérabilité ouvre de nombreuses perspectives.

Une ontologie coûte cher.

Oui, créer une ontologie peut être coûteux. Une ontologie est un composant logiciel et à ce titre on peut être amené à l'inclure dans un cycle de gestion de projet logiciel. Mais ce qu'il est important de reconnaître avant tout c'est qu'une ontologie n'est pas nécessairement très volumineuse ni nécessairement très formelle. La couverture, la spécificité, la granularité et la formalité d'une ontologie sont des caractéristiques variables d'une ontologie à l'autre en fonction des buts visés et du domaine d'application. D'autre part, du fait du caractère ouvert des ontologies, il est également possible d'aborder un domaine avec une ontologie minimaliste puis ensuite de l'étendre en fonction des besoins applicatifs et de modélisation.

Comme le disait Chris Welty (Conférence invitée à ISWC 2007) il ne s'agit pas tant du « web SEMANTIQUE » que du « WEB sémantique ». Ce qui est important dans le « web sémantique » c'est... le web. Et comme aime à le répéter Jim Hendler, à l'échelle du web un petit peu de sémantique peut déjà avoir des retombées gigantesques²⁸. Des schémas comme Dublin Core²⁹, FOAF³⁰ ou Creative Commons³¹ sont des preuves de ce qu'une sémantique petite mais virale peut accomplir et des exemples de ce qui est à notre portée et que nous devrions encourager, par exemple avec des initiatives comme les VoCamp³² où le côté social de la construction d'ontologie est également mis en avant.

De la même façon une ontologie, même petite, utilisée pour capturer et publier une folksonomie permet de multiplier les utilisations et réutilisations de ces métadonnées, comme nous le montrerons ci-après. Enfin les folksonomies ne sont pas gratuites, elles imposent d'attirer et de conserver un certain nombre d'utilisateurs actifs avant de voir une sémantique émerger de l'utilisation d'un système³³.

Qui dit ontologie, dit moteur d'inférence

²⁶ Euzenat J., Shvaiko P., *Ontology matching*, Springer-Verlag, Heidelberg (DE), 333p., 2007

<http://oaei.ontologymatching.org/>

²⁷ http://www.shirky.com/writings/ontology_overnated.html et <http://www.mindswap.org/blog/2007/11/21/shirkyng-my-responsibility/>

²⁸ <http://www.cs.rpi.edu/~hendler/LittleSemanticsWeb.html>

²⁹ <http://dublincore.org>

³⁰ Brickley D., *loc. cit.*

³¹ <http://creativecommons.org>

³² <http://vocamp.org>

³³ Halpin H., *loc. cit.*

Votre agenda électronique est une application du web sémantique dès lors qu'il est capable d'importer et d'exporter ses données et leur ontologie en utilisant les formalismes du web sémantique. Nous ne lui demanderons pas de faire autre chose que ce pour quoi il a été conçu (notifier, vérifier les conflits, etc.); par contre nous lui demandons de faire la seule chose qu'aucune autre application peut faire pour lui lorsqu'il s'agit de s'intégrer à d'autres applications : rendre ses schémas et ses données accessibles, expliciter son ontologie.

Une application à base d'ontologies n'est pas forcément une application qui intègre un moteur d'inférence. Un calendrier qui importe et exporte ses données au format du Web sémantique est une application du Web sémantique: il joue le jeu de l'interopérabilité et permet à d'autres applications non initialement identifiées de consommer ou compléter ses données ex: un blog, un service d'agence de voyage, un annonceur de spectacles, etc. Comme le rappelle le W3C. "Le Web Sémantique est un Web de Données"³⁴, une idée mise en application à grande échelle par l'initiative Linking Open Data³⁵ insistant sur la valeur offerte par la publication de données interconnectées et de leurs schémas suivant des formalismes standards.

Enfin revenons sur la notion d'inférence. Outre le fait que l'inférence logique ne soit pas un but en soi du Web sémantique, nous sommes aussi en droit de nous demander à quel point la recherche, la navigation, l'indexation, l'association, etc. ne reposent pas sur des inférences. Personnellement, nous ne réduisons pas l'inférence au sens latin (*infere*) de la causalité logique mais nous l'envisageons au sens large de raisonnement sur les circonstances où inférer peut être, par exemple, synonyme d'impliquer, indiquer, suggérer ou deviner. Nous utilisons les représentations des connaissances du Web sémantique dans bien d'autres manipulations formelles que la dérivation logique³⁶: approximation, comparaison, statistiques, clustering³⁷, etc. et dans des applications où la prise en compte de la dimension sociale et de l'utilisateur sont indéniables.

Les ontologies se gravent dans le marbre

Représenter et publier une ontologie par exemple en RDFS ou OWL ne signifie pas la graver dans le marbre. Le Web sémantique reconnaît parfaitement l'existence d'un cycle vie et y travaille à plusieurs titres (travaux sur les évolutions, relations d'équivalences entre différentes versions avec OWL, etc.) par exemple en discutant les bonnes pratiques d'utilisation des URI pour gérer l'évolution d'une notion en incluant un marqueur chronologique dans son identificateur. A titre d'exemple voici l'URI d'une primitive de SKOS (label préféré d'un concept) dans lequel on voit très nettement l'année et le mois du schéma la définissant :

<http://www.w3.org/2004/02/skos/core#prefLabel>

Le problème de choisir un nom (URI) fait par ailleurs l'objet de bonnes pratiques³⁸.

³⁴ <http://www.w3.org/2001/sw/>

³⁵ Bizer C., Heath T., Berners-Lee T., *Linked Data - The Story So Far*. International Journal on Semantic Web and Information Systems (IJSWIS), 5(3):1–22, 2009.

³⁶ Gandon F., *Graphes RDF et leur Manipulation pour la Gestion de Connaissances*, Habilitation à Diriger les Recherches (HDR), 5 novembre 2008, Inria, Université Nice - Sophia Antipolis

³⁷ Gandon F., Corby O., Giboin A., Gronnier N., Guigard C., *Graph-based inferences in a Semantic Web Server for the Cartography of Competencies in a Telecom Valley*, ISWC, Lecture Notes in Computer Science, Galway, 2005

³⁸ <http://www.w3.org/TR/cooluris/>

D'autre part, les ontologies sur le Web Sémantique sont également évolutives, et qui plus est de manière distribuée. Ainsi, il est possible à quiconque de définir une nouvelle classe, sous classe d'une classe d'une ontologie existante, si cette ontologie source ne propose pas de modélisation suffisamment précise.

Les ontologies arrivent trop tard

Dernière idée reçue : le web sémantique est voué à l'échec car il faudrait ré-analyser et transformer tous les documents du web existant, les représenter à l'aide d'ontologies, alors que les folksonomies grossissent avec les applications web qui les génèrent.

Le Web Sémantique est une extension du Web et non un remplaçant, comme le définit l'article fondateur du domaine³⁹. De plus il y a deux points intéressants ici : (1) une partie du web actuel est générée dynamiquement par des requêtes à des bases de données dont les données et les schémas pourraient être exposés sans nécessiter d'autre traitement (c'est notamment ce sur quoi travaille actuellement le groupe de travail RDF2RDB⁴⁰ au W3C (2) un certain nombre de données sont enterrées et dormantes dans des documents seulement parce qu'au moment où elles étaient explicites (ex : leur saisie dans un formulaire) nous n'avons pas su les capturer avec leurs schémas et les préserver dans un format accessible aux machines. Il s'agit alors de savoir capturer les données et leur schéma au moment où elles sont explicites afin de permettre des utilisations ultérieures et souvent même non prévues.

Un cas particulier ici est celui du tag: très souvent nous savons plus de choses sur le tag que ce que nous n'en représentons dans les folksonomies ex. la date, l'heure, la langue, le type de ressource taguée et l'application utilisée. Fournir des ontologies permettant de représenter ces actions de tagging de manière structurée permet ainsi de bénéficier de données existantes et de mettre celles-ci à disposition selon les formalismes du Web Sémantique, comme nous allons le discuter ci-après.

Ceci permet également d'établir un pont entre ce que l'on appelle souvent approche ascendante (folksonomies) et descendantes (ontologies). Ainsi, nous avons ici affaire à des "ontologies ascendantes" puisque contrôlées par les actions utilisateurs, ce qui remet en cause la vision, également fautive, que les ontologies sont forcément construites via des approches descendantes et contrôlées par des groupes restreints d'experts. Depuis le milieu des années 90 au moins trois types d'approches sont connus pour la construction d'ontologies (ascendantes, descendantes, centrifuges⁴¹) auxquelles s'ajoutent d'autres approches atypiques comme celle de l'élagage (SENSUS⁴²).

Après avoir discuté d'un point de vue théorique des caractéristiques et relations entre folksonomies, thésaurus et ontologies, nous allons maintenant adopter une deuxième forme d'argumentaire: la démonstration par l'exemple.

³⁹ Berners-Lee T., Hendler J.A., Lassila O., *The Semantic Web. Scientific American*, 284(5):34–43, 2001

⁴⁰ <http://www.w3.org/2001/sw/rdb2rdf/>

⁴¹ Uschold, M, Gruninger M, *Ontologies: Principles, methods and applications*, Knowledge Engineering Review 11(2), 1996, voir aussi AIAI-TR-191 AIAI, University of Edinburgh.

⁴² Swartout B., Ramesh P., Knight K., Russ T., *Toward Distributed Use of Large-Scale Ontologies*, AAAI Symposium on Ontological Engineering, Stanford, 1997

Initiatives de combinaison

Dans cette troisième partie, nous souhaitons construire sur le constat que l'opposition des ontologies, thésaurus et folksonomies comme des alternatives incompatibles est au mieux contre-productive et qu'au contraire leur intégration et leurs articulations recèlent de fortes valeurs ajoutées. Pour cela nous allons recenser les initiatives combinant ontologies, thésaurus et folksonomies et les classer selon trois familles d'approches:

- d'une part, les initiatives où la sémantique émergente au sein des folksonomies (cooccurrence, popularité, etc.) permet la construction (semi-)automatique d'ontologies ou de thésaurus (suggestion de structure, représentation intermédiaire, etc.).
- d'autre part, les initiatives où l'ontologie vient en model des folksonomies comme moyen de représenter les différents éléments entrant en jeu (tags, ressources, utilisateurs, applications, groupes, nuages, etc.) de manière structurée et interopérable.
- enfin les applications où deux ou trois de ces objets cohabitent.

Ontologies et thésaurus émergent des folksonomies

De nombreux travaux poursuivent l'objectif d'extraire des modèles structurés – taxonomies ou ontologies – depuis les folksonomies, principalement dans l'objectif de résoudre les problèmes classiques des systèmes à base de tags. L'objectif est alors d'explicitier la sémantique qui peut exister dans ces systèmes là où celle-ci n'est qu'implicite en raison de la nature même des folksonomies. La plupart d'entre eux se basent sur la notion de sémantique émergente⁴³ où l'usage collectif fait apparaître une sémantique contrôlée par la base (approche bottom-up) en opposition aux approches où la sémantique est définie en amont (top-down). Peter Mika évoque à ce sujet l'idée que les ontologies ainsi construites le seraient par émergence et non par un contrat rédigé par la majorité⁴⁴. L'ontologie (ou la taxonomie) émerge ainsi par effet de bord de l'architecture participative des systèmes à base de tags. Ce processus permet également de diminuer le goulot d'étranglement lié à l'acquisition d'un modèle structuré (étape généralement coûteuse) puisque ce modèle est ici issu des actions utilisateurs et des tags utilisés.

Afin d'identifier cette sémantique émergente au sein des folksonomies⁴⁵, propose ainsi une approche sociale de constitution d'ontologies. Il définit alors l'ontologie comme un modèle tripartite basé sur celui des folksonomies et ne considère plus uniquement les notions de classes et d'instances mais fait intervenir une composante sociale pour établir un modèle entre concepts (qu'il considère ici comme étant les tags), instances (les contenus tagués) et acteurs (les responsables des actions de tagging). Halpin *et al.*⁴⁶ se basent quant à eux sur une approche de cooccurrences réciproques entre tags pour extraire des relations taxonomiques, modélisées en RDFS, à partir d'une étude des bookmarks annotés par différents utilisateurs sur Delicious. Schmitz⁴⁷ propose également une approche basée sur les cooccurrences de tags et

⁴³ Staab S., *Emergent semantics*, IEEE Intelligent Systems, 17(1):78–86, 2002

⁴⁴ Mika P., *loc. cit.*

⁴⁵ *Ibid.*

⁴⁶ Halpin H., *loc. cit.*

⁴⁷ Schmitz P., *Inducing Ontology from Flickr Tags*. Actes WWW2006 Workshop on Collaborative Tagging, 2006

sur un modèle statistique de subsumption⁴⁸ pour établir une hiérarchie de tags depuis Flickr. Tout comme les deux approches précédentes, la sémantique des relations se résume à une unique relation de subsumption, permettant ainsi de passer de la folksonomie au thésaurus sans effort additionnel de la part de l'utilisateur. Cattuto *et al.*⁴⁹ comparent différentes mesures de similarité utilisées pour restructurer des ensembles de tags. Ils distinguent les mesures basées sur la cooccurrence des tags entre eux des mesures qui prennent en compte trois types différents de contextes d'association des tags : les tags associés à chacun des deux tags comparés, les ressources associées, et les utilisateurs associés.

La méthodologie FLOR (Folksonomy Ontology enRichment⁵⁰) définit quant à elle une méthode totalement non-supervisée (se basant notamment sur les résultats obtenus par Specia et Motta⁵¹) permettant d'explicitier la sémantique des tags et surtout des relations entre tags. Contrairement aux travaux précédents qui se limitent à des relations taxonomiques, leur approche permet d'extraire des relations typées entre concepts. Cette méthodologie repose notamment sur des notions de filtrage linguistique et d'expansion de termes et utilise différents outils proposés par le moteur sémantique Watson⁵². Avec FolksOntology⁵³ propose une approche semblable, l'utilisateur ayant en plus la possibilité de définir explicitement la sémantique des tags pour lesquels le système n'a pu trouver d'ontologie adaptée, i.e. de spécifier s'il s'agit d'une classe, d'une instance ou d'une propriété. On peut ainsi, plus qu'aligner la folksonomie avec des ontologies existantes, créer de nouveaux concepts.

Plusieurs approches aussi s'intéressent à l'application des distances d'édition entre termes aux tags (ex. distance de Levenshtein⁵⁴) afin de proposer des relations du type des thésaurus. Specia et Motta⁵⁵ ont appliqué cette approche à une folksonomie extraite de delicious.com conjointement à un alignement avec le thésaurus Wordnet afin de restructurer des tags.

Limpens compare⁵⁶ et combine⁵⁷ plusieurs des approches précédentes dans un cycle de vie d'un thésaurus construit et maintenu socialement.

Enfin d'un point de vue méthodologique les folksonomies peuvent s'ajouter aux côtés des thésaurus, des index et des lexiques à la liste des représentations intermédiaires d'une ontologie (approche MethOntology⁵⁸) i.e. aux documents plus ou moins structurés qui peuvent

⁴⁸ Sanderson M., Croft W.B., *Deriving concept hierarchies from text*. Actes Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR'99, pages 206–213. ACM Press, 1999

⁴⁹ Cattuto C., Benz D., Hotho A., Stumme G., *Semantic grounding of tag relatedness in social bookmarking systems*. 7th International Semantic Web Conference, 2008

⁵⁰ Angeletou S., *Semantic Enrichment of Folksonomy Tagspaces*, Actes International Semantic Web Conference, volume 5318 of Lecture Notes in Computer Science, pages 889–894. Springer, 2008 <http://flor.kmi.open.ac.uk>

⁵¹ Specia L., Motta E., *Integrating Folksonomies with the Semantic Web*, Actes European Semantic Web Conference, volume 4519 of Lecture Notes in Computer Science, pages 624–639. Springer, 2007

⁵² d'Aquin M., Sabou M., Motta E., Angeletou S., Gridinoc L., Lopez V., Zablith F., *What Can be Done with the Semantic Web? An Overview Watson-based Applications*. Actes Workshop on Semantic Web Applications and Perspectives, volume 426 of CEUR Workshop Proceedings. CEUR-WS.org, 2008.

⁵³ Van Damme C., *loc. cit.*

⁵⁴ Levenshtein V. I., *Binary codes capable of correcting deletions, insertions and reversals*, Soviet Physics Doklady., 10(8) , 1966, 707–710

⁵⁵ Specia L., Motta E., *loc. cit.*

⁵⁶ Limpens F., *loc. cit.*

⁵⁷ Limpens F., Gandon F., Buffa M., *Sémantique des folksonomies: structuration collaborative et assistée*, Conférence IC 2009, Hammamet, Tunisie, PUG, p37-48, 2009

⁵⁸ Fernández-López M, *loc. cit.*

être exploités en entrée des méthodes et outils de construction d'ontologies. Dans le cas particulier des approches linguistiques et terminologique de la construction des ontologies les tags avec leurs propriétés peuvent-être vus comme un nouveau type de candidats termes⁵⁹ aux côtés de ceux issus de l'analyse linguistique et en entrée d'outils de normalisation linguistique.

Ontologies des folksonomies

Comme nous l'avons évoqué auparavant, il est tout à fait possible d'utiliser des représentations ontologiques pour modéliser les folksonomies et leurs différents objets. Pour ce faire, Gruber⁶⁰ propose un premier modèle étendant la notion tripartite classique d'une action de tagging et où il définit celle-ci comme une relation faisant intervenir quatre éléments :

- un Objet, i.e. la ressource annotée quelque soit son type (billet de blog, photo, etc.) ;
- un Tag, i.e. le tag annotant la ressource ;
- un Agent, i.e. l'agent – en règle générale une personne – qui crée la relation ;
- une Source, i.e. l'espace où est effectuée cette action (e.g. Flickr).

C'est cette dernière propriété qui enrichit la relation initiale et qui permet de distinguer deux actions de tagging d'un même auteur pour la même ressource mais sur deux espaces distincts. Un cinquième élément peut également intervenir dans cette relation, à savoir une polarité permettant d'assigner une valeur positive ou négative à la relation, dans l'objectif de résoudre des problèmes de spam. Gruber⁶¹ introduit également la notion d'identité d'un tag et considère que des tags peuvent être définis comme identiques malgré des labels distincts, établissant un premier pas vers l'unification de tags hétérogènes et la notion de sens associés aux tags. Ce modèle théorique sera à la base du projet TagCommons⁶² qui ne proposera malheureusement pas d'ontologie RDFS/OWL prête à utiliser.

Newman *et al.*⁶³ définissent la Tag Ontology, une ontologie reprenant certains des principes définis par Gruber⁶⁴. Cette ontologie définit une classe tag:Tag pour modéliser les tags, sous-classe de skos:Concept, chaque tag disposant d'un ou plusieurs labels (via la propriété tag:tagName). En conséquence, il est possible d'établir des assertions RDF entre tags et notamment de les organiser en créant des relations de proximité entre eux. Le modèle définit ainsi une propriété tag:relatedTag (sous-propriété de skos:semanticRelation) pour représenter les relations possibles entre différents tags. Malheureusement, cette relation-ci ne porte pas suffisamment de sémantique pour définir si deux tags sont liés par proximité terminologique (ex. un tag est le pluriel d'un autre) ou parce qu'ils évoquent des domaines plus ou moins proches (ex. véhicule et voiture). Ce modèle définit également la notion d'action de tagging (tag:Tagging) et des relations à partir de celle-ci vers l'utilisateur (tag:taggedBy), le tag (tag:taggedWithTag) et la ressource associée (tag:taggedResource). Si la source et la polarité

⁵⁹ Aussenac-Gilles N., *loc. cit.*

⁶⁰ Gruber T. R., *Ontology of Folksonomy: A Mash-up of Apples and Oranges*. International Journal on Semantic Web and Information Systems, 3(2):1–11, 2007

⁶¹ Gruber T. R., *loc. cit.*

⁶² <http://tagcommons.org>

⁶³ Newman R., Ayers D., Russell S., *Tag ontology*, 2005. <http://www.holygoat.co.uk/owl/redwood/0.1/tags/>

⁶⁴ Gruber T. R., *loc. cit.*

de chaque action ne sont pas prises en compte dans ce modèle contrairement à Gruber⁶⁵ une composante temporelle peut-être ajoutée via la propriété tag:taggedOn.

Knerr⁶⁶ propose TagOnt qui reprend le modèle de Newman *et al.*⁶⁷ en y ajoutant la notion de visibilité d'une action de tagging, mais semble n'être utilisé dans aucun projet. Un modèle similaire est proposé par Echarte *et al.*⁶⁸ mais ne semble également pas avoir été utilisé en pratique.

Partant d'un besoin d'interopérabilité des tags entre applications, SCOT (*Semantic Cloud Of Tags*⁶⁹) se base sur les travaux précédents pour définir un modèle permettant l'export de l'ensemble des tags d'un utilisateur et leur fréquence d'utilisation d'un service vers un autre. Pour ce faire, SCOT introduit différentes classes et propriétés pour modéliser entre autres les cooccurrences entre tags au sein d'un système particulier (propriété scot:coocurlin et classe scot:Cooccurrence). Le modèle permet également de représenter plus finement que dans la Tag Ontology les relations entre tags, avec une dizaine de propriétés distinctes comme scot:acronym ou scot:plural.

Le Project NEPOMUK⁷⁰ propose via le vocabulaire NAO (NEPOMUK Annotation Ontology⁷¹) une classe nao:Tag et une propriété nao:has_tag pour identifier les tags rattachés à une ressource quelconque, sans pour autant considérer l'action de tagging en tant que modèle tripartite.

SIOC (Semantically-Interlinked Online Communities⁷²) quant à lui définit une simple classe Tag qui peut être utilisée en complément avec sioc:topic pour représenter les tags associés à un item. Il est également possible d'utiliser SKOS pour représenter des tags via la classe skos:Concept (sioc:Tag hérite d'ailleurs de cette classe), les instances associées pouvant ensuite être associées aux contenus tagués via la propriété sioc:topic.

Le modèle Bookmark⁷³, bien qu'il ne fasse pas explicitement référence à la notion de tags telle que popularisée ces dernières années, permet de faire le lien entre une ressource et un ensemble de termes annotant, représentés via une classe bookmark:Topic et une propriété bookmark:Bookmark. Il permet également d'organiser hiérarchiquement différentes instances de bookmark:Topic avec une propriété bookmark:subTopicOf, proposant une structuration similaire à la propriété skos:broader définie dans SKOS.

⁶⁵ Gruber T. R., *loc. cit.*

⁶⁶ Knerr T., *Tagging Ontology - Towards a Common Ontology for Folksonomies*, 2006, <http://code.google.com/p/tagont/>

⁶⁷ Newman R., *loc. cit.*

⁶⁸ Echarte F., Astrain JJ, Cordoba A., Villadangos J., *Ontology of Folksonomy: A New Modeling Method*, Actes Semantic Authoring, Annotation and Knowledge Markup Workshop, volume 289 of CEUR Workshop Proceedings. CEUR-WS.org, 2007

⁶⁹ Kim HL, Yang SK, Breslin JG, Kim HG., *Simple algorithms for representing tag frequencies in the scot exporter*. Actes IEEE/WIC/ACM International Conference on Intelligent Agent Technology, pages 536–539. IEEE Computer Society, 2007

⁷⁰ <http://nepomuk.semanticdesktop.org>

⁷¹ Scerri S., Sintek M., van Elst L., Handschuh S., *NEPOMUK Annotation Ontology Specification*, Technical report, 2007. <http://www.semanticdesktop.org/ontologies/nao/>

⁷² Breslin JG, Harth A., Bojars U., Decker S., *Towards Semantically-Interlinked Online Communities*. Actes European Semantic Web Conference, volume 3532 of Lecture Notes in Computer Science, pages 500–514. Springer, 2005, <http://sioc-project.org>

⁷³ Koivunen M.R., Swick R., Kaha J., Prud'hommeaux E., *An Annotea Bookmark Schema*. Technical report, World Wide Web Consortium, 2001. <http://www.w3.org/2003/07/Annotea/BookmarkSchema-20030707>

Le module Taxonomy de RSS 1.0⁷⁴ propose une propriété pour représenter les différents sujets associés à un élément de flux RSS.

Notons aussi l'existence du microformat rel:tag. Même s'il ne s'agit pas d'un modèle RDFS/OWL, l'utilisation de GRDDL⁷⁵, permet d'obtenir une représentation RDF des données annotées avec celui-ci.

Afin de franchir le pas entre folksonomies et ontologies MOAT (Meaning Of A Tag⁷⁶) propose une ontologie pour représenter la signification des tags (de manière générale et dans un contexte d'annotation particulier), cette signification étant représentée via une ressource termino-ontologique existante (un thésaurus existant, une ontologie existante, etc.). Ce modèle propose ainsi l'intégration des folksonomies dans le Web de données, en créant une passerelle entre le terme (tag) et le concept représenté (ressource du Web de données)⁷⁷. Pour ce faire, MOAT se base sur la Tag Ontology et introduit une classe Meaning pour représenter la signification d'un tag de manière globale (i.e. hors contexte), et une propriété tagMeaning pour associer cette signification dans le contexte d'une action particulière.

Enfin dernièrement, le modèle Nice Tag^{78,79,80} part du constat que les modélisations des tags dont nous disposons actuellement ne prennent pas suffisamment en considération leur richesse et leur diversité. NiceTag propose une ontologie dans laquelle les tags sont assimilés à des graphes nommés qui peuvent être structurés, typés et annotés autant que nécessaire pour capturer au maximum l'action de tagging, son contexte et son résultat. Les tags 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.). Chaque élément de ce triplet (la ressource, le signe et le lien) peut être typé et caractérisé autant que nécessaire. 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.

Applications logicielles combinant les deux approches

⁷⁴ <http://web.resource.org/rss/1.0/modules/taxonomy/>

⁷⁵ GRDDL Use Cases: Scenarios of extracting RDF data from XML documents. W3C Working Group Note 6 April 2007, World Wide Web Consortium, 2007. <http://www.w3.org/TR/grddl-scenarios/>

⁷⁶ Passant A., Laublet P., *Meaning Of A Tag: A collaborative approach to bridge the gap between tagging and Linked Data*. Actes WWW2008 Workshop Linked Data on the Web, volume 369 of CEUR Workshop Proceedings. CEUR-WS.org, 2008, <http://moat-project.org>

⁷⁷ Passant A., Laublet P., Breslin JG, Decker S., *A URI is Worth a Thousand Tags: From Tagging to Linked Data with MOAT*, International Journal on Semantic Web and Information Systems (IJSWIS), 5(3):71–94, 2009

⁷⁸ Monnin A., Limpens F., Gandon F., Laniado D., *Speech acts meets tagging: NiceTag ontology*, AIS SigPrag International Pragmatic Web Conference Track, I-Semantics, Graz, Austria, September 2010

⁷⁹ Limpens F., Monnin A., Laniado D., Gandon F., *NiceTag Ontology: tags as named graphs*, International Workshop on Social Networks Interoperability at ASWC09

⁸⁰ Monnin A., Limpens F., Laniado D., Gandon F., *L'ontologie NiceTag : les tags en tant que graphes nommés*, Atelier Web Social, 10^{ème} Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances, EGC, Hammamet, Tunisia, 2010

Utilisant certains des modèles précédents, différents outils permettent de faire le lien entre folksonomies, thésaurus et ontologies. Ainsi, dès 2001, Annotea⁸¹ propose un système d'annotations et de partage de ressources Web ouvert et reposant sur les technologies du web sémantique, proposant les annotations en RDF selon son propre vocabulaire⁸². On peut ainsi considérer Annotea comme une des premières applications sociales de partage de contenus basé sur les technologies du Semantic Web.

Dans cette même idée de représenter des contenus annotés avec les technologies du web sémantique, Revyu⁸³ est un service de revues utilisant la Tag Ontology pour permettre une représentation de ses tags et actions de tagging. Faviki⁸⁴ offre un service de gestion de favoris où les tags sont des identifiants DBpedia (un export RDF de Wikipédia⁸⁵). Il prend ainsi en compte la notion de multilinguisme associée aux tags, puisqu'une même URI peut être associée à plusieurs termes. De la même manière LODr⁸⁶ offre cette possibilité d'indexer des tags existant depuis des services Web 2.0 (Flickr, Twitter, etc.) à des identifiants DBpedia pour rendre la recherche de contenus personnels plus pertinente, puisque basée sur des concepts et non plus uniquement sur des termes.

D'autres outils sont axés plus spécifiquement sur la gestion des tags, et sur la manière de les organiser pour pallier à leurs limites. Ainsi, les outils de bookmarking WebMark^{87,88}, SemanticScuttle⁸⁹, Gnizr⁹⁰ et Semanlink⁹¹ permettent de définir des relations hiérarchiques entre tags, le second offrant un export RDF des contenus annotés en utilisant les modèles précédents (Tag Ontology, SIOC et SKOS), le dernier étant basé sur son propre modèle de représentation des tags reposant sur SKOS. Dans une approche différente, GroupMe⁹² propose aux utilisateurs de regrouper les tags par catégories pour faciliter la recherche d'information, représentant le tout avec sa propre ontologie⁹³. Sweetwiki⁹⁴ permet également l'organisation de

⁸¹ Kahan J., Koivunen M-R., *Annotea: an open RDF infrastructure for shared Web annotations*. International World Wide Web Conference (WWW10), pages 623–632, 2001 <http://www.w3.org/2001/Annotea/>

⁸² <http://www.w3.org/2000/10/annotation-ns#>

⁸³ Heath T., Motta E., *Revyu.com: A Reviewing and Rating Site for the Web of Data*, International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), volume 4825 of Lecture Notes in Computer Science, pages 895–902. Springer, 2007, <http://revyu.com>

⁸⁴ Milicic V., *Semantic tags*, W3C SWEO Case Study, World Wide Web Consortium, 2008. <http://www.w3.org/2001/sw/sweo/public/UseCases/Faviki/> et <http://faviki.com>

⁸⁵ Auer S., Bizer C., Lehmann J., Kobilarov G., Cyganiak R., Ives Z., *Dbpedia: A nucleus for a web of open data*, International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), volume 4825 of Lecture Notes in Computer Science, pages 715–728. Springer, 2007, <http://dbpedia.org>

⁸⁶ Passant A., *Linked Data tagging with LODr*, Semantic Web Challenge 2008, collocated with the 7th International Semantic Web Conference (ISWC), 2007, <http://lodr.info>

⁸⁷ Delaforge N., Gandon F., Vukosavljevic N., *Webmark: pour une approche contextuelle de la référence en ligne*, Atelier Philosophie et Ingénierie, Le formel face à l'histoire, la technologie et la matérialité, IC2011, Chambéry, France, 2011

⁸⁸ Delaforge N., Gandon F., *Webmarks : Le marquage d'intérêt sur le Web de données*, Proc. 12e Conférence Internationale Francophone sur l'Extraction et la Gestion de Connaissance, EGC, Bordeaux, 2012

⁸⁹ Huynh-Kim-Bang B., Dané E., *Social bookmarking et tags structurés*. In IC2008, 19e Journées Francophones d'Ingénierie des Connaissances, 2008

⁹⁰ <http://code.google.com/p/gnizr>

⁹¹ Servant FP, *Semanlink*. Jena User Conference (JUC), 2006

⁹² <http://groupme.org>

⁹³ Abel F., Frank M., Henze N., Krause D., Plappert D., Siehndel P., *GroupMe! – Where Semantic Web Meets Web 2.0*, International Semantic Web Conference and 2nd Asian Semantic Web Conference (ISWC/ASWC2007), volume 4825 of Lecture Notes in Computer Science, pages 871–878. Springer, 2007

tags (et utilise son propre modèle de représentation), cette fois-ci au sein d'un wiki. L'approche reste fidèle à la philosophie wiki⁹⁵ en permettant à tous les utilisateurs du système de gérer cette organisation commune de l'ensemble des tags du wiki de manière ouverte et collaborative.

S'ils n'utilisent pas explicitement les technologies du web sémantique, d'autres outils permettent manuellement de structurer ou d'enrichir les systèmes à base de tags et de bénéficier de ces enrichissements au moment de la recherche d'information, en passant notamment de systèmes de tags à des thésaurus. Ainsi, toujours dans une approche permettant de dériver des relations taxonomiques à partir de folksonomies, Jäschke *et al.*⁹⁶ proposent aux utilisateurs de Bibsonomy⁹⁷ (outil collaboratif de gestion de références bibliographiques) de définir eux-mêmes des relations hiérarchiques entre tags. Tanasescu et Streibel⁹⁸ proposent une autre solution pour structurer les folksonomies avec l'Extreme Tagging. Les utilisateurs ont ici la possibilité de typer les tags et les relations entre ces tags, à nouveau en utilisant des tags. Enfin, on peut également citer les machine tags de Flickr⁹⁹, où les utilisateurs peuvent définir des tags prédicat=objet, par exemple dct:description=New-York ou geo:lat=42.33. S'ils ne sont nativement pas modélisés en RDF, ces machine tags peuvent être traduits comme tels via l'API Flickr¹⁰⁰.

Conclusion

Il est impossible de prévoir tous les usages de nos données mais il est possible de les permettre si nous les préservons et si nous les ouvrons. C'est là la raison d'être du web de données.

L'intelligence qui émergerait d'un web sémantique n'est pas tant dans une connexion d'intelligences que dans une intelligente connexion. Le choix d'être socio-conscient ou sémio-conscient est un choix d'opérationnalisation et non une contrainte des standards ou des modèles. Les ontologies en général, et le Web sémantique en particulier, ne s'opposent pas aux dimensions sociales, sémiotiques ou pragmatiques des scénarios d'usage; dans le pire des cas ce sont des opérationnalisations particulières qui peuvent être incriminées mais jamais les standards.

Nous pensons que des scénarios complexes et des solutions complètes pourront articuler des phases reposant sur différentes structures (folksonomie, thésaurus, ontologie) au sein d'une

⁹⁴ Buffa M., Gandon F.L., Ereteo G., Sander P., Faron C., *SweetWiki: A semantic wiki*, Journal of Web Semantics, 6(1):84–97, 2008

⁹⁵ Krötzsch M., Schaffert S., Vrandečić D., *Reasoning in semantic wikis*. In Grigoris Antoniou, Uwe Amann, Cristina Baroglio, Stefan Decker, Nicola Henze, Paula-Lavinia Patranjan, and Robert Tolksdorf, editors, Reasoning Web, volume 4636 of Lecture Notes in Computer Science, pages 310–329. Springer, 2007

⁹⁶ Jäschke R., Hotho A., Schmitz C., Ganter B., Stumme G., *Discovering Shared Conceptualizations in Folksonomies*, Journal of Web Semantics, 6(1):38–53, 2008.

⁹⁷ <http://bibsonomy.org>

⁹⁸ Tanasescu V., Streibel O., *Extreme Tagging: Emergent Semantics through the Tagging of Tags*, International Workshop on Emergent Semantics and Ontology Evolution (ESOE2007), volume 292 of CEUR Workshop Proceedings. CEUR-WS.org, 2007

⁹⁹ <http://www.flickr.com/groups/mtags/>

¹⁰⁰ <http://librdf.org/flickrurl/>

même application, d'autant plus que, comme nous l'avons montré, différents ponts existent entre ces artefacts cognitifs en théorie et en pratique. Ainsi, il est tout à fait possible de partir d'un aspect social et participatif (folksonomies) à partir duquel des structures plus formelles, offrant des capacités avancées de raisonnement, peuvent être extraites (thésaurus et ontologies) ou combinées. C'est de cette manière que nous pourrions parvenir à un Web à la fois Social et Sémantique, rejoignant les premières idées de Berners-Lee et des "machines sociales"¹⁰¹.

Ainsi pour amener le web et ses ressources à leur plein potentiel, il est important de ne pas opposer les modèles existants de classification, d'organisation ou d'indexation et les pratiques émergent¹⁰² du web mais bien de rechercher activement leur complémentarité et leur interaction pour que ces artefacts s'enrichissent mutuellement dans des cycles de vie^{103,104} et d'utilisation couplés avec, à la clef, la diversité, la richesse et l'échelle du web.

¹⁰¹ Berners-Lee T., *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its Inventor*. Harper Collins Publishers, New York, 1999

¹⁰² Gandon F., Giboin A., *Vers des ontologies à l'état sauvage*, Atelier IC 2.0, joint aux IC2008, 19èmes Journées Francophones d'Ingénierie des Connaissances, Nancy, 2008

¹⁰³ Limpens F., Gandon F., Buffa M., *A complete life-cycle for the semantic enrichment of folksonomies*, Advances In Knowledge Discovery and Management - Vol 3, F. Guillet, P. Pinaud, G. Venturini and D. Zighed (eds), 2013, p133 Springer

¹⁰⁴ Limpens F., Gandon F., Buffa M., *Un cycle de vie complet pour l'enrichissement sémantique des folksonomies*, Proc. 11eme Conférence Internationale Francophone sur l'Extraction et la Gestion des Connaissances, EGC 2011