

Combiner Web 2.0 et Web Sémantique pour réduire les disparités d'expertise au sein de blogs d'entreprise

Alexandre Passant*, Philippe Laublet**

*Digital Enterprise Research Institute, National University of Ireland, Galway
IDA Business Park, Lower Dangan, Ireland
alexandre.passant@deri.org,

**LaLIC, Université Paris-IV Sorbonne
Maison de la Recherche, 28 rue Serpente, 75006 Paris, France
philippe.laublet@paris-sorbonne.fr

Résumé. Avec l'avènement d'applications sociales en entreprise (blogs, wikis, etc.), il est fréquent que des individus aux niveaux d'expertise relativement distants se réunissent au sein de communautés en ligne. Ces disparités d'expertise se traduisent entre autres par des comportements différents dans la manière de tagguer les contenus créés, notamment en ce qui concerne les termes utilisés, rendant ainsi complexe la découverte d'informations pourtant publiées. Dans cet article, nous mettons en avant la possibilité offerte par les technologies du Web Sémantique, combinées avec les paradigmes du Web Social, de résoudre cette problématique. Nous proposons ainsi une chaîne de traitement combinant ontologies, wikis sémantiques et indexation de contenus permettant la production de graphes sémantiques interconnectés et facilitant de cette manière la découverte de contenus créés au sein de tels systèmes¹.

1 Introduction

Alors que les outils comme les blogs et les wikis ainsi que les paradigmes sociaux associés sont aujourd'hui couramment acceptés sur le Web, sous l'appellation commune Web 2.0, ces pratiques font depuis quelques années leur apparition en entreprise, au sein de ce que l'on nomme généralement Entreprise 2.0 (McAfee, 2006). Parmi les usages introduits par cette mouvance, nous notons l'utilisation abondante du *tagging*, association de mots-clés libres aux contenus publiés dans un objectif de classification contrôlée par l'utilisateur². Or, l'hétérogénéité des niveaux d'expertise représentés au sein de ces communautés (composées aussi bien d'experts que de stagiaires) se traduit par l'utilisation de tags plus ou moins spécifiques pour annoter les contenus produits. Nous avons ainsi pu constater ces disparités dans un contexte d'entreprise (Passant, 2009), comme d'autres ont pu auparavant l'identifier sur le Web (Golder et Huberman, 2006). Partant de ce constat, nous proposons ici une solution basée sur les

1. Une partie des travaux présentés dans cet article à été réalisée au sein d'EDF R&D (bourse CIFRE) et est actuellement financée par la Science Foundation Ireland au sein du projet Lion 2 (SFI/08/CE/I1380).

2. <http://www.vanderwal.net/folksonomy.html>

principes et technologies du Web Sémantique combinées aux pratiques et outils du Web 2.0 pour répondre à cette problématique. En particulier, notre proposition repose sur (1) l'utilisation d'ontologies légères pour représenter la sémantique des tags et (2) la prise en compte d'interactions sociales pour créer et maintenir les bases de connaissances associées. Cette combinaison permet ainsi de produire un ensemble unifié d'annotations sémantiques liant experts, non experts et contenus produits et annotés.

2 Représenter la sémantique des tags avec MOAT et l'utilisation de wikis sémantiques

Alors que Web 2.0 et Web Sémantique ont longtemps été considérés, à tort, comme disjoints, différents travaux récents ont montré que la combinaison de ces deux approches était doublement bénéfique (Gandon, 2006) (Gruber, 2007). Plus précisément, différentes contributions ont été proposées ces dernières années dans le but de combiner folksonomies et ontologies, recensées notamment par (Limpens et al., 2009), aussi bien en terme de modèles ontologiques pour représenter les tags que de systèmes permettant de dériver des taxonomies ou ontologies de domaine depuis des folksonomies. Nous concernant, nous avons fait le choix d'une approche qui consiste à utiliser des ressources formelles identifiées par leur URI (particulièrement des instances de classes d'ontologies de domaine) en support des tags de manière à représenter la signification de ceux-ci, le tout modélisé avec une ontologie spécifique, nommée MOAT — Meaning Of A Tag³ (Passant et al., 2009). Ce modèle, combiné à d'autres comme la *Tag Ontology*⁴, FOAF — Friend of a Friend⁵ — et SIOC — Semantically-Interlinked Online Communities⁶ — permet ainsi de représenter sans ambiguïté le sens donné à un tag dans un contexte d'annotation particulier. Par exemple il est possible de modéliser que la signification du tag TF (acronyme de *Thin film*, un type particulier de cellule solaire) correspond à la ressource identifiée par l'URI `example:ThinFilm` (Fig. 1). Si cette proposition permet de résoudre les problèmes d'hétérogénéité et d'ambiguïté des tags (un autre tag `thinfilm` pouvant être associé au même concept), elle permet également de résoudre le manque d'organisation de ceux-ci. En effet, via l'utilisation d'ontologies de domaine, nous bénéficions non seulement d'identifiants universels et non ambigus pour représenter ces significations mais également de relations entre les ressources associées, qui permettent ainsi de lier contenus et instances d'ontologies de manière uniforme et globale comme nous le verrons dans la section suivante. En complément de ce modèle, nous avons également défini un *framework* logiciel permettant aux utilisateurs d'associer leurs tags à de telles URIs, en utilisant à nouveau des principes collaboratifs pour partager et faire évoluer ces relations au sein de communautés en ligne⁷.

Si le modèle précédent permet de lier tags et instances de classes d'ontologies de domaine, la question se pose de savoir comment représenter et maintenir de telles instances. Si l'on peut utiliser des bases de connaissances publiques, notamment celles proposées dans le cadre du

3. <http://moat-project.org>

4. <http://www.holygoat.co.uk/projects/tags/>

5. <http://foaf-project.org>

6. <http://sioc-project.org>

7. <http://moat-project.org/architecture>

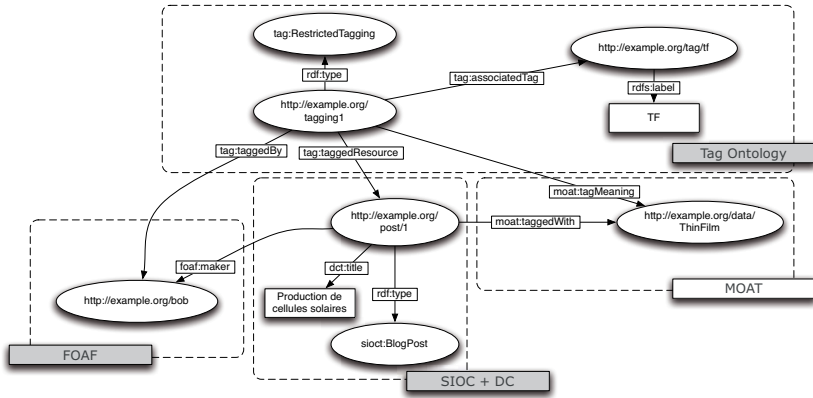


FIG. 1 – Indexation de contenus taggués avec MOAT

project Linking Open Data⁸, ce qui permet notamment d'utiliser MOAT non seulement en entreprise mais plus généralement sur le Web ouvert, il est plus judicieux dans notre contexte de se référer à des ressources internes à l'entreprise. En effet, un des bénéfices de notre modèle étant l'utilisation des relations entre ressources pour découvrir de nouveaux contenus, utiliser des ressources internes permet d'enrichir ces relations et de s'assurer de la validité et de la consistance des ressources utilisées. Dans ce cadre, nous proposons l'utilisation de wikis sémantiques dans le but de créer et maintenir de telles instances, et ce de manière collaborative en bénéficiant de la philosophie wiki (Schaffert, 2006).

Nous avons ainsi implémenté un prototype particulier de wiki sémantique, permettant de définir des *templates* de pages associés à des ontologies de domaines et taxonomies internes, représentées en RDFS/OWL. Ce prototype a permis la création de plus de 350 instances maintenues par une vingtaine d'utilisateurs⁹, les instances ainsi créées étant par la suite utilisées pour associer les significations aux tags via MOAT, l'association se faisant manuellement avec cependant un système de suggestion et un aspect collaboratif permettant de partager celles-ci via le *framework* précédent. Nous bénéficions donc d'une chaîne de traitement complète pour enrichir notre folksonomie et la plate-forme de blogs associée en liant billets de blogs et instances d'ontologies interconnectées, en combinant MOAT et wiki sémantiques.

3 Bénéfices pour la recherche d'information

En se basant sur les annotations produites via notre chaîne de traitement, la figure qui suit (Fig. 2) montre de quelle manière deux billets de blog sont interconnectés à partir du moment où l'un a été associé, avec MOAT, à `<http://example.org/data/ThinFilm>` et le second à `<http://example.org/data/Solaire>`, ces deux instances étant liées par une relation `skos:broader` modélisée via notre wiki sémantique (Fig. 2).

8. <http://linkeddata.org>

9. Analyse sur 200 jours.

Disparités d'expertise au sein de blogs d'entreprise

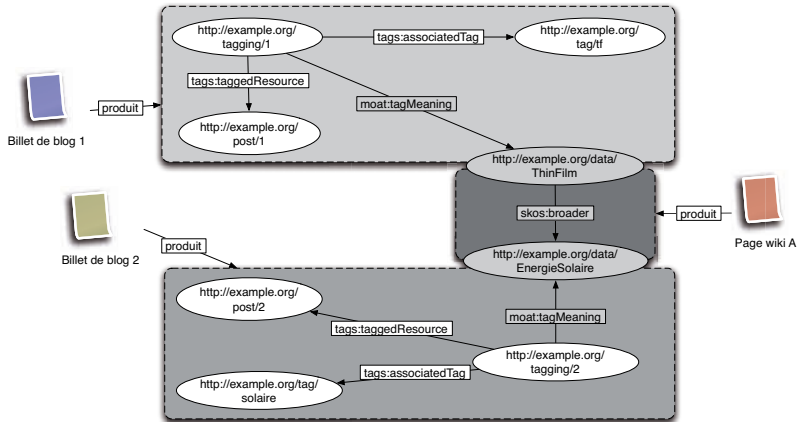


FIG. 2 – Identification de contenus proches via des relations entre concepts associés

Afin de bénéficier de ces relations, nous avons mis en place un moteur de recherche sémantique qui, entre autres, identifie pour une ressource donnée, l'ensemble des billets de blogs annotés par cette ressource (la recherche se faisant donc au niveau de la ressource et non plus du simple mot-clé, résolvant les problèmes initiaux d'hétérogénéité et d'ambiguïté). De plus, ce moteur est capable de suggérer des concepts proches, ce qui permet de prendre en compte les problèmes d'hétérogénéité d'expertise et ainsi de lister un certain nombre de domaines spécifique lors d'une recherche générique. Pour ce faire, nous avons mis en place un système de suggestion de concepts proches en définissant pour différentes classes de nos ontologies, des règles d'inférence, comme par exemple (Listing 1) qui identifie que deux entités sont considérées comme proches à partir du moment où l'une évolue dans un domaine plus spécifique que l'autre, cette relation de spécificité étant modélisée avec SKOS — Simple Knowledge Organization System¹⁰. Si ces règles sont définies formellement, dans notre exemple en N3 (mais nous pourrions également imaginer utiliser RIF — Rules Interchange Format¹¹ — ou OWL2¹² dans certains cas), celles-ci sont implémentées sous forme de requêtes SPARQL¹³ pour des raisons de maturité des outils associés. En terme d'interface, chaque concept proche ainsi identifié est proposé sous forme de lien hypertexte vers la page associée au sein du moteur de recherche, afin d'accéder aux documents annotés correspondants.

Ces principes de suggestion de concepts proches nous permettent ainsi de prendre en compte le problème des différents niveaux d'expertise évoqué motivé initialement. En effet, nous avons identifié dans notre contexte que les non-experts avaient tendance à utiliser des tags représentant des concepts de haut niveau (e.g. solaire) là où les experts utilisaient des tags beaucoup plus spécifiques (e.g. TF pour *Thin Film*). L'apport de ces règles d'inférence permet donc d'établir une passerelle entre les concepts génériques et les concepts spécifiques,

10. <http://www.w3.org/2004/02/skos/>

11. <http://www.w3.org/TR/rif-core/>

12. <http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>

13. <http://www.w3.org/TR/rdf-sparql-query/>

```

{
  xxx a role:Domain .
  yyy a role:Domain .
  xxx skos:broaderTransitive yyy .
} => {
  xxx :related yyy .
}

```

Listing 1 – Règle d'inférence basée sur SKOS pour l'identification de concepts proches

et par extension entre les contenus annotés par des non-experts et les contenus annotés par des experts, comme le montre la figure qui suit (Fig. 3).

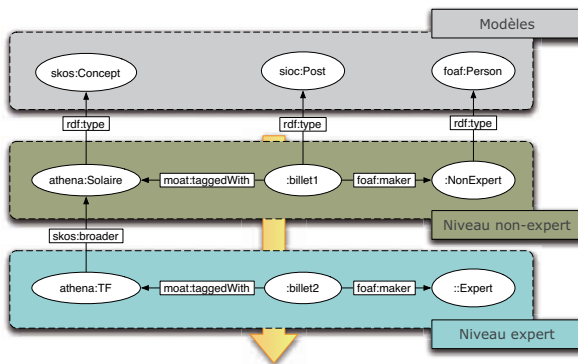


FIG. 3 – Relations entre experts et non-experts en combinant FOAF, SIOC, MOAT et SKOS

En analysant d'autre part notre folksonomie d'origine et les différentes annotations RDF représentées avec MOAT, nous avons constaté que 1176 tags avaient été associés à 715 URIs de significations différentes. Nous n'avons constaté que très peu de tags sujets aux problématiques d'ambiguïté dans notre contexte, seul un d'entre eux étant associé à plusieurs URIs¹⁴. En contrepartie, nous avons constaté un problème d'hétérogénéité beaucoup plus présent. Nous avons ainsi observé que si 510 URIs sur les 715 recensées ne sont pas sujettes à des problèmes d'hétérogénéité, puisqu'assignées à un seul tags, 205 le sont. 96 instances ont ainsi été associées à deux tags, 70 à trois d'entre eux et 39 à quatre tags ou plus. Nous avons également observé que si cette hétérogénéité est en général le fait de plusieurs utilisateurs annotant avec différents tags, elle peut également émerger à un niveau personnel. Nous avons plus particulièrement constaté cette hétérogénéité personnelle au niveau de tags et d'instances représentatifs de noms de personne (nom complet et nom de famille), de zones géographiques (par exemple USA et états-unis et de technologies (synonymie, multilinguisme mais aussi abréviations).

14. Notons que cela ne signifie pas qu'un seul tag est ambigu dans la folksonomie, puisque seuls 1176 tags sur un total de 12257 ont été ici considérés.

4 Conclusion

Dans cet article, nous avons proposé une méthodologie combinant Web Social et Web Sémantique afin de produire un ensemble de connaissances formelles (instances d'ontologies de domaine) venant en support de systèmes de *tagging* existants. Nous avons ainsi détaillé l'utilisation couplée d'un modèle, MOAT, permettant d'associer tags et ressources du Web Sémantique et d'un système de wiki sémantique afin d'offrir une chaîne complète de production d'annotations sémantiques en support de systèmes existants. Nous avons également montré de quelle manière nous avons mis en place un tel processus, en revenant de plus sur la manière dont notre proposition permet d'établir un graphe complet d'annotations sémantiques liant des contenus annotés avec des tags reflétant différents niveaux d'expertise.

Références

- Gandon, F. L. (2006). Le web sémantique n'est pas antisocial. In *IC2006, 17èmes Journées Francophones d'Ingénierie des Connaissances*, pp. 131–140.
- Golder, S. et B. A. Huberman (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science* 32(2), 198–208.
- Gruber, T. (2007). Ontology of Folksonomy : A Mash-up of Apples and Oranges. *International Journal on Semantic Web and Information Systems* 3(2), 1–11.
- Limpens, F., F. Gandon, et M. Buffa (2009). Linking Folksonomies and Ontologies for Supporting Knowledge Sharing : a State of the Art. Délivrable ISICIL T3.5.
- Mcafee, A. P. (2006). Enterprise 2.0 : The Dawn of Emergent Collaboration. *MIT Sloan Management Review* 47(3), 21–28.
- Passant, A. (2009). *Technologies du Web Sémantique pour l'Entreprise 2.0*. Ph. D. thesis, Université Paris-Sorbonne.
- Passant, A., P. Laublet, J. Breslin, et S. Decker (2009). A URI is Worth a Thousand Tags : From Tagging to Linked Data with MOAT. *International Journal on Semantic Web and Information Systems* 5(3), 1–11.
- Schaffert, S. (2006). IkeWiki : A Semantic Wiki for Collaborative Knowledge Management. In *Workshop on Semantic Technologies in Collaborative Applications (STICA 06)*.

Summary

Due to the growing number of social applications in enterprise contexts (blogs, wikis, etc.) people with heterogeneous expertise level often gather together in online communities. This heterogeneity of expertise generally leads to different behaviors regarding information tagging, especially regarding the terms used, making information retrieval context, while the content to be retrieved is online. In this paper, we describe the capabilities offered by Semantic Web technologies, combined with Social Web paradigms, to solve this issue. We then propose a food chain combining ontologies, semantic wikis and content indexation in order to provide interlinked semantic graphs, facilitating the discovery of content created within these ecosystems.