

# **ISICIL : Intégration Sémantique d'Informations à travers des Communautés d'Intelligence en Ligne**

Pavel Arapov, Sébastien Comos, Olivier Corby, Nicolas Delaforge, Guillaume Erétéo, Catherine Faron Zucker, Michel Buffa, Fabien Gandon, Guillaume Husson, Freddy Limpens

INRIA, 2004 route des Lucioles, BP93, 06902 Sophia Antipolis,  
I3S – Université de Nice Sophia Antipolis – CNRS – UMR 6070 - 930 route des Colles,  
BP145, 06903 Sophia Antipolis  
[isicil-contact@lists-sop.inria.fr](mailto:isicil-contact@lists-sop.inria.fr)  
<http://isicil.inria.fr>

## **1 Introduction**

Le défi du projet ISICIL est de concilier les nouvelles applications virales du web avec des représentations formelles et des processus d'entreprise pour les intégrer dans les pratiques de veille en entreprise. Nous concevons, étudions et expérimentons de nouveaux outils d'assistance aux tâches d'intelligence en entreprise (voir [Leitzelman, Bugeaud et Al] pour un état de l'art sur ce sujet). Ces outils s'appuient sur les interfaces avancées des applications du web 2.0 (blog, wiki, social bookmarking) pour les interactions et sur les technologies du web sémantique pour l'interopérabilité et le traitement de l'information. ISICIL permet aussi d'explorer les nouveaux développements scientifiques de la notion de collaboration épistémique (interactions humaines orientées vers la transmission et l'élaboration de savoirs) et de repérer des axes technologiques exploitables dans les prolongements de ce projet. Les prototypes développés dans le cadre d'ISICIL vont être déployés auprès des personnes en charge de la veille technologique chez l'ADEME et Orange Labs.

## **2 Conception des algorithmes et modèles de la plate-forme**

Pour la partie serveur web sémantique, le moteur sémantique CORESE/KGRAM est interfacé avec un serveur d'application J2EE et une base de données qui assurent la modularité et l'accès sécurisé aux données et métadonnées. Dans cette première moitié du projet, plusieurs modules ont été développés : un service de gestion des tags générés par les utilisateurs, un service de gestion des utilisateurs, un service de web scraping et un wiki sémantique.

### **2.1 Serveur de tags**

Nous avons développé une première méthode semi-automatique pour la structuration sémantique de la folksonomie (ensemble des tags) : elle consiste à comparer et à sélectionner une trentaine de distances syntaxiques pour le calcul des variations orthographiques entre tags, que nous avons complété avec des métriques d'usage (co-utilisation, inclusion de communautés). L'utilisation de SKOS (description de thésaurus), SIOC et SCOT (pour les tags) pour modéliser le résultat de l'analyse a été étendue par de nouvelles ontologies publiées et libres (SRTag et NiceTag) qui supportent les contradictions sémantiques. Pour capturer le retour utilisateur sur la structuration des folksonomies, ces modèles sont intégrés

à un gestionnaire de bookmarks : sur une recherche l'interface propose les variations calculées et l'utilisateur peut accepter, refuser ou ajuster ces propositions ; ses préférences sont capturées et ajoutées au modèle. L'exploitation des retours utilisateur est en cours notamment en utilisant des règles de résolution de conflits et des opinions d'utilisateurs référents. [Limpens et al.]

## 2.2 Serveur de profils

Nous avons conçu et implanté une série d'opérateurs d'analyse sémantique de réseaux sociaux et de schémas d'analyse. Une convention de recherche signée avec le réseau social Ipernity nous a permis de tester les algorithmes sur un réseau de 60 000 membres et 500 000 relations typées transformé en RDF. Une première version du modèle utilisé pour la représentation des communautés a été testée ainsi que plusieurs formalisations SPARQL pour le calcul d'indicateurs de centralité paramétrés par une ontologie. Une campagne de tests complète est en cours ainsi qu'une extension à l'identification de communautés. [Erétéo et al.]

## 2.3 Scraping et wiki sémantique

Le web scraping consiste à effectuer une copie locale d'un document web ou d'une partie d'un document web, afin de l'archiver, de le partager, de le commenter, de l'annoter. Nous avons couplé ce service avec le wiki sémantique (pour le stockage du document copié), avec un serveur de ressources (pour les métadonnées) et avec un système de gestion des droits d'accès utilisé par l'ensemble des outils ISICIL [Buffa et Faron]

## 2.4 Interfaces intégrées à la navigation

Pour la partie cliente, Firefox et son système d'extensions ont été choisis notamment pour la récupération et l'intégration des contenus existants : l'idée majeure consiste à intervenir directement dans les modèles des solutions CMS/GED/Forum/Blog/Intranet en place chez les utilisateurs afin d'enrichir les contenus publiés par des métadonnées (RDFa). Une première extension Firefox détecte les données d'une page en cours de consultation, afin d'augmenter la navigation. Une deuxième extension Firefox permet d'effectuer l'opération de scraping avec archivage sur le wiki de la plate-forme. Lors de cette opération, il est possible d'ajouter une description au document et de le taguer. L'extension génère également du RDF décrivant ce document.

## 3 Démonstration

Cette proposition de démonstration présentera les modules listés ici et les résultats obtenus sur plusieurs jeux de données.

## Bibliographie complémentaire

Buffa, M., Faron-Zucker,(2010), C., *Gestion sémantique des droits d'accès au contenu de sites web: l'ontologie AMO*, soumis à la 10ème Conférence Internationale Francophone

- "Extraction et Gestion des Connaissances" EGC'10, Hammamet, Tunisie, 26-29 Janvier 2010.
- Bugeaud, F., Soulier, E., (2010), *Méthodes et modèles pour les Systèmes d'Information*. In: *Master système d'information*, coord. Thévenot J., (Collection Master, dirigée par J. Igalens), Paris, Eska.
- Bugeaud, F., Giboin, A., Soulier, E., (2010), *Shared Representations for Innovation: Experience Feedback on Two Innovation-oriented Projects*, COOP'10, Workshop The mediation role of shared representations in cooperative activities: new challenges.
- Ereteo, G., Gandon, F., Corby, O., Buffa, M. (2009), *Analysis of a Real Online Social Network Using Semantic Web Frameworks*, In Proceedings of the 8th International Semantic Web Conference, ISWC2009.
- Ereteo, G., Gandon, F., Corby, O., Buffa, M. (2009), *Semantic Social Network Analysis*, In Proceeding of the first Web Science Conference, WebSci'09.
- Leitzelman, M., (2010), *La veille 2.0 : comment outiller les interactions sociales au sein du processus de veille*. Dans : *Les Cahiers du Numérique - Du web 2.0 au concept 2.0* sous la direction de : Luc Quoniam, édition Hermes
- Limpens, F., (2010), *Multi-points of View Semantic Enrichment of Folksonomies*, PhD thesis, University of Nice - Sophia Antipolis, <http://hal.archives-ouvertes.fr/hal-00530714/fr>
- Limpens, F., Gandon, F., Buffa, M., (2010), *Helping online communities to semantically enrich folksonomies*, In Proceedings of the WebSci10: Extending the Frontiers of Society On-Line, April 26-27th, 2010, Raleigh, NC: US.
- Limpens, F., Monnin, A., Laniado, D., Gandon, F., (2009), *NiceTag Ontology: tags as named graphs*, International Workshop in Social Networks Interoperability, ASWC09.
- Freddy Limpens, Fabien Gandon, Michel Buffa, (2009), *"Collaborative semantic structuring of folksonomies"*, (short paper), IEEE/WIC/ACM Int. Conf. on Web Intelligence.
- Gandon, F., Corby, O., (2010) *Name That Graph, or the need to provide a model and syntax extension to specify the provenance of RDF graphs*, W3C Workshop RDF next step

## Summary

On the one hand, Web 2.0 tools exhibit limits when it comes to automating some tasks or controlling some processes, as usually required in a corporate environment. On the other hand, more structured information systems often suffer from usability and knowledge capture issues. In addition, in the context of intelligence, corporate structures can also provide assistance at different stages of these processes to ensure that corporate quality standards and rules are met. The challenge of the ISICIL project is to reconcile viral new web applications with formal representations and processes to integrate them into corporate practices for technological and scientific monitoring. More specifically, ISICIL proposes to study and to experiment with the usage of new tools for assisting corporate intelligence tasks. These tools rely on web 2.0 advanced interfaces (blog, wiki, social bookmarking) for interactions and on semantic web technologies for interoperability and information processing.