

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
<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