

Un cycle de vie complet pour l'enrichissement sémantique des folksonomies

Freddy Limpens*, Fabien Gandon*
Michel Buffa**

*EPI Edelweiss - INRIA Sophia Antipolis
Freddy.Limpens, Fabien.Gandon@inria.fr,
<http://www-sop.inria.fr/edelweiss/>

**I3S - KEWI, UNSA - CNRS
buffa@unice.fr
<http://www.i3s.unice.fr/I3S/labos/lab02.html>

Résumé. Les tags fournis par les utilisateurs des plateformes de tagging social ne sont pas explicitement liés sémantiquement, et ceci limite considérablement les possibilités d'exploitation de ces données. Nous présentons dans cet article notre approche pour l'enrichissement sémantique des folksonomies qui intègre une combinaison de traitements automatiques ainsi que la capture des contributions de structuration des utilisateurs via une interface ergonomique. De plus, notre modèle supporte les points de vue qui divergent tout en permettant de les combiner en respectant leur cohérence locale. Cette approche s'adresse aux communautés de connaissances collaborant en ligne, et en intégrant leurs usages, nous sommes en mesure de proposer un cycle de vie complet pour le processus de structuration sémantique des folksonomies. La navigation dans les données de tagging est ainsi améliorée, et les folksonomies peuvent alors être directement intégrées dans la construction de thesauri.

1 Introduction

Le tagging social est un moyen simple et populaire de classification par mot clé. L'ensemble des mots utilisés forme une folksonomie que l'on peut considérer comme "le vocabulaire de la communauté". Cependant les folksonomies résultant de cette pratique ne possèdent aucune structure car les tags ne sont pas reliés sémantiquement et manquent de précision, notamment à cause des variations orthographiques, ce qui limite leur potentiels d'exploitation pour la navigation dans les corpus tagués. Nous présentons dans cet article notre approche de l'enrichissement sémantique de folksonomie. Notre contribution consiste en un cycle complet du processus qui combine traitements automatiques et contributions des utilisateurs.

Un exemple-type de nos communautés cibles est représenté par l'Ademe¹ qui cherche à améliorer la structuration et l'indexation des sources d'information utilisées dans les processus de veille scientifique et technique. Dans ce scénario, nous pouvons distinguer trois types

1. Agence De l'Environnement et de la Maitrise de l'Energie <http://www.ademe.fr>