

Extraction de relations pour le peuplement d'une base de connaissance à partir de tweets

Cédric Lopez*, Elena Cabrio**, Frédérique Segond*

*Viseo R&D, 4, avenue Doyen Louis Weil, Grenoble, France
{cedric.lopez, frederique.segond}@viseo.com
<http://www.viseo.com/fr/offre/recherche-et-innovation>

**Université Côte d'Azur, Inria, CNRS, I3S, France
elena.cabrio@unice.fr
<http://wimmics.inria.fr/>

Résumé. Dans une base de connaissance, les entités se veulent pérennes mais certains événements induisent que les relations entre ces entités sont instables. C'est notamment le cas pour des relations entre organisations, produits, ou marques, entités qui peuvent être rachetées. Dans cet article, nous proposons une approche permettant d'extraire des relations d'appartenance entre deux entités afin de peupler une base de connaissance. L'extraction des relations à partir d'une source dynamique d'informations telle que Twitter permet d'atteindre cet objectif en temps réel. L'approche consiste à modéliser les événements en s'appuyant sur une ressource lexico-sémantique. Une fois les entités liées au Web des données ouvertes (en particulier DBpedia), des règles linguistiques sont appliquées pour finalement générer les triplets RDF qui représentent les événements.

1 Introduction

La structuration automatique de données textuelles brutes est une tâche particulièrement importante pour aider à la construction du Web sémantique. Récemment, nous avons développé *SMILK plugin*, un outil qui permet de structurer des données acquises au cours de la navigation d'un utilisateur sur des pages Web afin de les stocker dans une base de connaissance au format du Web sémantique (*triplestore*) (Lopez et al., 2016). Or, les triplets générés sont composés d'entités nommées telles que des noms de produits, de marques, ou d'organisations, et ont une durée de validité limitée. Par exemple, le triplet "Yves Saint Laurent Beauté, appartient au groupe, Pinault-Printemps-Redoute" est obsolète depuis le rachat de cette organisation par le groupe L'Oréal, en 2008. La base de connaissance devrait donc être remise à jour avec le nouveau triplet "Yves Saint Laurent Beauté, appartient au groupe, L'Oréal". La possibilité qu'un changement puisse survenir à tout moment ne permet pas d'assurer la pérennité des triplets de la base de connaissance ce qui peut conduire à un raisonnement erroné. Par exemple, le rachat d'une marque appartenant à une société S1 par une autre société S2 implique que toutes les gammes, produits, *etc.* de cette marque deviennent obsolètes pour la société S1, ce qui implique un résultat faux lorsque l'on traite des requêtes telles que "Quel est l'impact sur