

Approche logique pour la réconciliation de références

Fatiha Saïs *, Nathalie Pernelle *
Marie-Christine Rousset**

* LRI, Université Paris-Sud 11, F-91405 Orsay Cedex,
INRIA Futurs, 2-4 rue Jacques Monod, F-91893 Orsay Cedex, France
{prenom.nom}@lri.fr

**LSR-IMAG BP 72, 38402 St MARTIN D'HERES CEDEX

Marie-Christine.Rousset@imag.fr

<http://www-lsr.imag.fr/Les.Personnes/Marie-Christine.Rousset/>

Résumé. Le problème de réconciliation de références consiste à décider si deux descriptions provenant de sources distinctes réfèrent ou non à la même entité du monde réel. Dans cet article, nous étudions ce problème quand le schéma des données est décrit en RDFS étendu par certaines primitives de OWL-DL. Nous décrivons et montrons l'intérêt d'une approche logique basée sur des règles de réconciliation qui peuvent être générées automatiquement à partir des axiomes du schéma. Ces règles traduisent de façon déclarative les dépendances entre réconciliations qui découlent de la sémantique du schéma. Les premiers résultats ont été obtenus sur des données réelles dans le cadre du projet PICSEL 3 en collaboration avec France Telecom R&D.

1 Introduction

Le problème de réconciliation de références est un problème majeur pour l'intégration ou la fusion de données provenant de plusieurs sources. Il consiste à décider si deux descriptions provenant de sources distinctes réfèrent ou non à la même entité du monde réel (e.g., la même personne, le même article, le même gène, le même hôtel).

Il est très difficile d'attaquer ce problème dans toute sa généralité car les causes d'hétérogénéité dans la description de données provenant de différentes sources sont variées et peuvent être de nature très différente. L'hétérogénéité des schémas est une des causes premières de la disparité de description des données entre sources. De nombreux travaux, dont on peut trouver une synthèse dans Rahm et Bernstein (2001); Shvaiko et Euzenat (2005); Noy (2004); Euzenat et Valtchev (2004), ont proposé des solutions pour réconcilier des schémas ou des ontologies par des mappings. Ces mappings peuvent ensuite être utilisés pour traduire des requêtes de l'interface de requêtes d'une source vers l'interface de requête d'une autre source.

L'homogénéité ou la réconciliation de schémas n'empêchent cependant pas les variations entre les descriptions des instances elles-mêmes. Par exemple, deux descriptions de personnes avec les mêmes attributs Nom, Prénom, Adresse peuvent différer sur certaines valeurs de ces attributs tout en référant à la même personne, par exemple, si dans l'un des tuples le prénom est en entier alors que dans l'autre tuple il n'est donné qu'en abrégé.