

Ré-écriture de requêtes dans un système d'intégration sémantique

Cheikh Niang^{*,**}, Béatrice Bouchou^{*}, Moussa Lo^{**}, Yacine Sam^{*}

^{*}Université François Rabelais Tours, Laboratoire d'Informatique
3 place Jean Jaurès, 41000 Blois — {prénom.nom}@univ-tours.fr

^{**}Université Gaston Berger, Laboratoire d'Analyse Numérique et d'Informatique
Route de Ngalléle, BP 234 Saint-louis, Sénégal — moussa.lo@ugb.edu.sn

Résumé. Nous décrivons la deuxième phase de réalisation d'un système d'intégration qui minimise l'intervention humaine habituellement nécessaire. Après la phase de construction semi-automatique du schéma (ontologie) global décrite dans de précédents articles, nous présentons ici le processus de ré-écriture de requêtes globales en des requêtes adressées aux sources.

1 Introduction

L'intégration de données vise à combiner des données issues de sources différentes afin de permettre leur exploitation à travers une interface d'interrogation appelée schéma global. Le travail que nous présentons ici est une des phases de réalisation d'un système d'intégration de type médiation (approche non matérialisée) sémantique. Ce dernier est un triplet $\mathcal{J} = \langle \mathcal{G}, \mathcal{S}, \mathcal{M} \rangle$ (Lenzerini (2002)) où, dans notre proposition, le schéma global \mathcal{G} et les schémas locaux \mathcal{S} sont décrits par des ontologies. Nous construisons le schéma global semi-automatiquement à partir des schémas sources, ce qui permet de définir naturellement un mapping GAV. Pour autant, comme montré dans Niang et al. (2011), notre construction incrémentale de \mathcal{G} induit que l'ajout et le retrait de sources ont un coût de mise à jour de \mathcal{G} très limité. Nous montrons ici qu'en plus cela n'a *aucun effet sur le mapping* que nous définissons, car ce mapping *n'est calculé qu'au moment où une requête doit être évaluée*, en fonction de l'état courant de \mathcal{G} et *uniquement pour ce qui concerne la requête*.

Notre contribution vise une intégration dynamique au sein de communautés web. Une source peut rejoindre le système intégré le temps d'échanger ses données ou de contribuer à une tâche particulière et peut se retirer à tout moment. Les collaborations qui se multiplient dans de nombreux domaines suscitent la constitution de solides ontologies de référence, publiquement accessibles. C'est dans ce cadre que nous avons mis au point dans Niang et al. (2011) notre construction du schéma global, en exploitant une ontologie de référence. Ce schéma global est une TBox spécifiée en *DL-Lite_A*, une logique de description (LD) (Calvanese et al. (2007)) connue pour traiter des requêtes dans un système d'intégration avec une complexité LOGSPACE en la taille des données interrogées. Nous exploitons dans cet article cette LD pour la **ré-écriture** d'une requête q , exprimée dans les termes de \mathcal{G} , en des sous-requêtes ciblant des sources susceptibles d'y répondre.