

Techniques de fouille de données pour la réécriture de requêtes en présence de contraintes de valeurs

Hélène Jaudoin*, Frédéric Flouvat*

*Laboratoire LIMOS, UMR CNRS 6158
Université Blaise Pascal - Clermont-Ferrand II,
63 177 Aubière cedex, France
{hjaudoin,flouvat}@isima.fr

Résumé. Dans cet article, nous montrons comment les techniques de fouilles de données peuvent résoudre efficacement le problème de la réécriture de requêtes en termes de vues en présence de contraintes de valeurs. A partir d'une formalisation du problème de la réécriture dans le cadre de la logique de description $\mathcal{ALN}(\mathcal{O}_v)$, nous montrons comment ce problème se rattache à un cadre de découverte de connaissances dans les bases de données. L'exploitation de ce cadre nous permet de bénéficier de solutions algorithmiques existantes pour la résolution du problème de réécriture. Nous proposons une implémentation de cette approche, puis nous l'expérimentons. Les premiers résultats démontrent l'intérêt d'une telle approche en termes de capacité à traiter un grand nombre de sources de données.

1 Introduction

Aujourd'hui, les techniques d'analyse et d'intégration de données sont devenues des atouts majeurs pour les entreprises et les services gouvernementaux. En effet, ces techniques permettent un gain de temps pour regrouper et croiser l'information distribuée. Dans le domaine du développement durable, ces techniques sont notamment indispensables afin de rassembler et d'analyser les pratiques agricoles et ainsi garantir la traçabilité des pratiques. Plus précisément nos travaux se situent dans le cadre d'un projet¹ visant à mettre en place un système d'intégration pour interroger les sources de données agricoles distribuées. Le système doit être flexible pour permettre l'arrivée de nouvelles sources de données afin de suivre le processus d'informatisation du domaine agricole. En effet entre 2000 et 2003, le nombre d'exploitations ayant un accès à Internet a triplé². Il doit de plus permettre de traiter un grand nombre de sources de données car le domaine est susceptible d'accueillir, en plus des services déconcentrés des ministères, un grand nombre d'exploitations.

Dans cet article, nous nous plaçons dans le cadre d'un système de médiation suivant une approche Local As View (LAV), où les vues sont décrites via des requêtes sur le schéma global. Cette approche est connue pour être flexible car l'ajout et la suppression de sources de

¹Ce projet est réalisé en collaboration avec le Cemagref, <http://www.cemagref.fr/>

²<http://www.acta-informatique.fr/>