

Visualisation par l'exemple des dépendances dans les bases de données relationnelles

Fabien De Marchi*, Jean-Marc Petit**

*Laboratoire LIRIS, UMR CNRS 5205

Université Claude Bernard - Lyon 1

8, boulevard Niels Bohr, 69 622 Villeurbanne cedex France

fabien.demarchi@liris.cnrs.fr

**Laboratoire LIMOS, UMR CNRS 6158

Université Blaise Pascal - Clermont-Ferrand II

24 avenue des Landais, 63 177 Aubière cedex, France

jmpetit@math.univ-bpclermont.fr

Résumé. Comprendre la sémantique des bases de données relationnelles existantes est important pour de nombreuses applications. Cette sémantique est principalement véhiculée par les dépendances fonctionnelles (DF) et les dépendances d'inclusion (DI) ; elles généralisent respectivement les notions de clé et de clé étrangère. Toutefois, il est fréquent que les bases de données opérationnelles deviennent désordonnées dans le temps ; dans ce cas, les contraintes d'intégrité doivent être retrouvées à partir des données. Plusieurs méthodes ont été proposées pour la découverte des DF ou des DI. Ces algorithmes fournissent à l'administrateur un ensemble de dépendances satisfaites dans les données.

Se pose alors le problème de la compréhension des dépendances extraites, incluant des aspects liés à la visualisation des connaissances. Cette étape doit permettre, par exemple, d'assister l'utilisateur final à sélectionner les règles intéressantes, ou à comprendre pourquoi une dépendance attendue n'est pas satisfaite dans les données. Nous proposons de fournir à l'administrateur ou l'analyste, en complément de la liste des règles, un échantillon de la base de données, vérifiant exactement les même DF et DI, appelé *base de données d'Armstrong informative* (BDAI). Ces exemples nous semblent particulièrement adaptés pour faciliter les échanges entre l'administrateur et les experts du domaine. Nous donnons certaines propriétés sur l'existence et la taille des BDAI, ainsi que des algorithmes pour les construire. Des expérimentations sur une base réelle issue du web montrent l'intérêt pratique de cette proposition.

1 Introduction

Comprendre la sémantique des bases de données relationnelles existantes est important pour de nombreuses applications. Parmi elles, citons des travaux de rétro-conceptions [Casanova et de Sa, 1983, Markowitz et Makowsky, 1990, Petit *et al.*, 1996, Comyn-Wattiau et Akoka, 1999], dont le but est de retrouver le schéma conceptuel des données à partir de leur forme relationnelle, des travaux sur l'intégration de données