

Une approche fondée sur la corrélation entre prédicats pour le traitement des réponses pléthoriques

Patrick Bosc*, Allel Hadjali*, Olivier Pivert*, Grégory Smits**

*Irisa ENSSAT - Univ. Rennes 1, Lannion France
{bosc,hadjali,pivert}@enssat.fr

**Irisa IUT Lannion - Univ. Rennes 1, Lannion, France
gregory.smits@univ-rennes1.fr

Résumé. L'interrogation de bases de données, dont les dimensions ne cessent de croître, se heurte fréquemment au problème de la gestion des réponses pléthoriques. Une des approches envisageables pour réduire l'ensemble des résultats retournés et le rendre exploitable est de contraindre la requête initiale par l'ajout de nouvelles conditions. L'approche présentée dans cet article s'appuie sur l'identification de liens de corrélation entre prédicats associés aux attributs de la relation concernée. La requête initiale peut ainsi être intensifiée automatiquement ou par validation de l'utilisateur à travers l'ajout de prédicats proches sémantiquement de ceux spécifiés.

1 Introduction

Depuis le début des années 90, il a été reconnu que l'enrichissement des systèmes d'accès à l'information par des approches coopératives était devenu un besoin pragmatique et indéniable. Gaasterland (1992) souligne que le principal enjeu des approches coopératives d'accès à l'information est de fournir des résultats corrects, fiables et utiles en réponse aux requêtes utilisateurs au lieu de résultats littéraux. Les réponses coopératives ont également pour objectif de fournir des résultats répondant aux besoins et attentes des utilisateurs et ne nécessitant aucun effort et traitement supplémentaires pour être examinés et exploités.

En ce sens, deux problèmes distincts sont fréquemment soulevés. Le premier concerne la gestion des requêtes retournant un ensemble de résultats vide. Lorsque aucun enregistrement ne répond pleinement à la requête soumise par l'utilisateur, l'enjeu des approches coopératives est alors de fournir des réponses alternatives. Le second problème concerne le cas des requêtes fournissant des résultats pléthoriques. Ces requêtes retournent un nombre de réponses trop important pour être exploitées efficacement et facilement par l'utilisateur. Cette situation problématique est d'autant plus actuelle que la présence et l'accès à des bases de données de plus en plus volumineuses se démocratisent et se généralisent, notamment via Internet. Parmi l'ensemble volumineux des résultats retournés, l'utilisateur n'est généralement intéressé que par un sous-ensemble de ces résultats qu'il juge plus pertinent. Pour extraire ce sous-ensemble, l'utilisateur doit parcourir la totalité des résultats retournés, ce qui constitue une tâche fastidieuse et malaisée. Cet article vise à résoudre ce dernier problème en proposant une méthode