

Gestion de données hétérogènes dans un entrepôt de données

Laurence Duval

Crest-Ensai
Campus de Ker Lann
35170 Bruz
laurence.duval@ensai.fr

L'un des problèmes majeurs lors de la constitution d'entrepôts de données est la consolidation des données issues des différentes sources. De nombreux travaux dans le domaine ont permis d'améliorer la qualité des données de l'entrepôt ([Calvanese et al. 2001], [Florescu et al., 2001], [Métais et Sèdes, 2002],...), mais il existe toujours des cas où la valeur retenue n'est pas certaine. Actuellement aucun entrepôt relationnel permet de stocker plusieurs valeurs pour une même donnée lorsque les informations issues des sources divergent. Cette limitation a des conséquences sur la qualité des données de l'entrepôt.

Nous proposons dans ces travaux de stocker dans une base de données objet-relationnelle des attributs objets auxquels peuvent être associées plusieurs valeurs. Ceci nous permet de conserver les valeurs de plusieurs sources pour une même donnée lorsque les valeurs des sources divergent. Nous pouvons constituer un entrepôt sans perdre les informations contenues dans les sources. L'utilisation d'un tel entrepôt peut être transparente pour l'utilisateur, dans ce cas il accède aux données qui seraient présentes dans un entrepôt usuel. Un langage a été défini pour permettre d'accéder aux nouvelles valeurs. Par exemple, l'accès aux données peut se faire en privilégiant une source pour l'ensemble des attributs de la requête ou pour un attribut particulier.

Nous nous appuyons sur la gestion des objets dans une base de données objet-relationnelle qui nous permet d'utiliser des attributs multi-valués tout en conservant l'algèbre relationnelle. Un prototype a été réalisé en utilisant le système de gestion de bases de données Oracle 9i.

Références

- [Calvanese et al. 2001] D. Calvanese, G. De Giacomo, M. Lenzerini, D. Nardi, R. Rosati, *Data integration in datawarehousing*, International Journal of Cooperative Information Systems, Vol. 10, 3: 237-271, 2001
- [Florescu et al., 2001] D. Florescu, D. Shasha, E. Simon, C.-A. Saita, *Declarative Data Cleaning: Language, Model and Algorithms*, Proceeding of the 27th VLDB Conference, Roma, Italie, 2001
- [Métais et Sèdes, 2002] E. Métais, F. Sèdes, *Appariement d'informations dans les entrepôts de données : quelques approches pour le filtrage flexible*, Information-Interaction-Intelligence, Volume 2, N°2, 2002