

Université Montpellier 2 - LIRMM

Encadrant : [Z. Bellahsène](#)

Intitulé : Structures d'index pour optimiser la recherche de mappings entre arbres XML

Mots clés : Intégration de données, intégration de schéma, XML, ontologie, index.

Présentation de la problématique

Le principe de médiation consiste à fournir à un utilisateur un accès uniforme et transparent à la localisation et l'hétérogénéité des données. Quelle que soit l'architecture adoptée, "entrepôt de données" ou "virtuelle", la définition d'un schéma médiateur est nécessaire pour jouer le rôle d'interface uniforme sur lequel l'utilisateur peut formuler ses requêtes. Pour cela, les relations de correspondance (ou mapping) entre les sources de données et le schéma médiateur doivent être spécifiées ou de préférence automatiquement générées afin que les requêtes exprimées sur le schéma médiateur puissent être traduites sur les schémas des sources. Nous nous intéressons ici à la recherche de correspondances (*mapping*) entre arbres dans le contexte XML.

Plusieurs méthodes de recherche de mappings ont été proposées, ce sujet de stage vise à étendre nos propres travaux pour prendre en compte l'existence d'index et d'ontologies de domaine pour optimiser la recherche de mappings dans un contexte de grande échelle où le nombre de schéma est très grand.

Travail à faire

1. Etude bibliographique du problème d'intégration de schémas
2. Etude bibliographique des structures d'index pour des arbres XML
3. Proposer des algorithmes de recherche de mapping qui reposent sur l'existence d'une part des ontologies de domaine et des structures index sur les schémas à intégrer.
4. Implémentation, mise en oeuvre et évaluation sur des exemples.

Références

O'Connor M., Bellahsène Z., Roantree M., "An Extended Preorder Index for Optimising XPath Expressions", XML Database Symposium, Springer Verlag, Trondheim, Norway, 2005, LNCS 3671, pp. 114- 128.

Jayant Madhavan Philip A. Bernstein AnHai Doan Alon Halevy, Corpus-based Schema Matching, International Conference on Data & Knowledge Engineering, IEEE, ICDE2005.

Tranier J., Baraër R., Bellahsène Z., Teisseire M.: Where's Charlie: Family-Based Heuristics for Peer-to-Peer Schema Integration. [IDEAS 2004](#): 227-235.