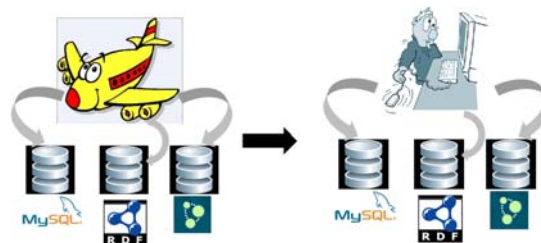


Grandes bases de connaissances : stockage et distribution

JF Baget (baget@lirmm.fr) & M Croitoru (croitoru@lirmm.fr)

Nous souhaitons réaliser une application aidant à la réservation des voyages. Toutes les informations liées aux vols (horaires, provenance, destination, prix, ...) sont stockées dans différentes bases de données relationnelles. Par contre, les informations liées à la proximité géographique entre différentes villes et leurs aéroports, comme celles liées à la situation géographique et au prix d'hôtels se trouvent dans divers triplestores RDF (<http://en.wikipedia.org/wiki/Triplestore>, http://fr.wikipedia.org/wiki/Resource_Description_Framework). Pour répondre à nos requêtes de façon encore plus pertinente, nous souhaitons utiliser une base de connaissance encodant les avis de consommateurs, que l'on peut trouver dans une base de données de graphes (<http://neo4j.org/>).

Il s'agit donc pour nous de disposer d'un mécanisme de requêtage distribué, permettant d'accéder à des éléments de connaissances stockées dans différents formats (nous ne nous intéresserons pas ici au problème lié à l'alignement d'ontologies). L'objectif de ce stage est de préparer la réalisation d'un tel moteur de raisonnement, en considérant les points suivants :



(1) Une première étape du travail concerne l'étude comparative entre différents modes de stockage existants : triplestores, bases relationnelles, graphes... Les paramètres de l'étude seront par exemple la taille nécessaire au stockage, le coût (en temps) de l'insertion ou de la suppression d'information...

(2) Une deuxième étape consiste à comparer la complexité des opérations élémentaires nécessaires à un algorithme de type backtrack (<http://en.wikipedia.org/wiki/Backtracking>) qui interrogerait une base écrite dans ces différents formats.

(3) Dans une dernière étape, il s'agira de distribuer l'algorithme de backtrack et d'implémenter des wrappers permettant d'interroger de façon uniforme les bases de connaissances, quel que soit le mode de stockage considéré.

Ce sujet peut être considéré comme une introduction à la proposition de thèse que vous pourrez trouver à l'adresse suivante : <http://www.lirmm.fr/~croitoru/these-JF-MC.pdf>.

