

Projet Forum

Compte-rendu de la réunion du Jeudi 23 mars 2006, Montpellier

Présents : Z. Bellahsene, S. Benbernou, P. Bosc, M. Bouet, M. Cart, J. Ferrié, W. Guedria, A. Hadjali, H. Jaudoin, H. Kherdouci, F. De Marchi, O. Pivert, V. Prince, F. Pinet, M. Roche, K. Saleem, F. Toumani, et F. Vigier.

Contexte. Grève des étudiants dans la plus part des universités y compris celle de Montpellier où certaines salles et issues étaient bloquées. Ceci a entraîné un retard pour le démarrage de la réunion.

10h15 Début de la réunion

Présentation du programme de la journée. L'objectif de la journée était de faire un état l'art sur les problématiques étudiées dans le projet FORUM.

10h30 Découverte de mappings entre schémas, Khalid Saleem, LIRMM

Définition du problème et présentation des différentes approches et outils existant pour effectuer la recherche de correspondances ou mappings entre des ontologies/schémas.

Actuellement, les techniques proposées se sont focalisées sur la qualité des correspondances avec l'intervention d'un expert et négligeant quelque peu le facteur de performance qui est aussi important notamment lorsque le nombre de schémas devient très grand.

Khalid Saleem a présenté l'esquisse d'une proposition basée sur la "fouille d'arbres" (tree mining) pour optimiser l'espace de recherche de mappings dans le contexte à large échelle ou le nombre de schémas devient très grand.

La discussion porte sur l'existence de langages pour décrire les mappings.

11h15 Utilisation de techniques de TAL pour le mapping des schémas, Mathieu Roche, LIRMM

Porte sur la normalisation des labels, et sur les différentes similarités entre des termes utilisés dans le traitement du langage. Outre les mesures de similarité entre termes fondés sur des mesures terminologiques, un stagiaire de Master Recherche travaille actuellement au LIRMM sur la prise en compte de contextes plus riches (labels des arbres, mots-clés, description en langage naturel, etc.) pour améliorer la qualité du mapping. Lors de la discussion, nous cherchons à savoir si nous pouvons nous appuyer sur une description riche des schémas et éventuellement des instances pour améliorer la qualité des mappings entre schémas. De telles données pourraient aider à la résolution des problèmes d'homonymie et de synonymie. Une collaboration avec le CEMAGREF devra donc être mise en oeuvre pour travailler dans ce sens.

12h Déjeuner

13h50 Requêtes flexibles et unification de représentations hétérogènes Patrick Bosc, IRISA

L'exposé a porté sur les différentes approches pour introduire des préférences dans les requêtes en commençant par présenter la différence entre un ensemble ordinaire et un ensemble flou à savoir la transition graduelle entre l'appartenance totale et l'exclusion. Par la suite, l'exposé s'est focalisé sur l'utilisation de la théorie des ensembles flous pour introduire des préférences dans les requêtes et représenter les requêtes flexibles et montrer par des

exemples le rôle des préférences dans les requêtes et l'extension d'opérateurs de l'algèbre relationnelle.

La discussion a porté sur les connexions possibles avec la découverte de mappings en plus des requêtes flexibles. Farouk propose d'utiliser la notion de préférence pour limiter le nombre de sources à interroger.

15h00 Réécriture de requêtes, Farouk Toumani, LIMOS

L'exposé portait sur l'intégration d'information et le problème du traitement des requêtes dans les différentes approches de médiation. La sémantique des réponses à une requête dans le cadre des systèmes d'intégration a été définie par Abiteboul et Dushka en termes de *réponses certaines*, i.e. réponses correctes quelque soit le contenu réel des sources de données. Une technique habituellement utilisée pour calculer ces réponses est celle de la réécriture de requêtes en utilisant les vues. Les problèmes soulevés par la réécriture ont été présentés dans le cadre des différentes approches de médiation GAV/LAV. Un cadre général de réécriture a été ensuite présenté : étant donnée une requête Q et un ensemble des vues V , le problème consiste à reformuler Q en une expression Q' , appelée réécriture, qui utilise uniquement les vues de V telle que $Q' \rho Q$ (ρ peut être la subsomption ou l'équivalence) et telle que Q' optimise un certain critère (par exemple un coût).

Dans l'optique de mise en place d'un système d'intégration flexible, la définition de la réécriture peut être relaxée selon différents points. La relation ρ devient une relation d'**approximation** au lieu des relations classiques de subsomption et d'équivalence. Le critère d'optimisation peut être étendu pour prendre en compte une notion de **qualité de service** ou de **préférence** utilisateur. D'autres directions peuvent être également explorées comme la réécriture de requêtes en présence de prédicats flous ou l'utilisation de la théorie des possibilités pour tenir compte de l'incertitude des données.

15h35 pause

16h00 Restructuration de graphes dynamiques, Hammamache Kheddouci, LIRIS

L'exposé a porté sur les changements topologiques de graphes dans le temps. Techniques propres au domaine des graphes, permettant de résonner sur des graphes qui évoluent dans leur structure. Les applications possibles sont : le routage, la découverte de services, etc. La discussion a porté sur l'application possible pour la découverte de mappings avec des techniques de plongements de graphes. Les algorithmes sont à construire.

16h35 Contexte applicatif et les problèmes qui se posent d'un point de vue hétérogénéité et interopérabilité, François Pinet, CEMAGREF

L'exposé a porté sur le contexte applicatif de FORUM : « le suivi de l'épandage agricole en France ». Il s'agit de gérer l'épandage et d'assurer la traçabilité qui s'est faite jusqu'ici manuellement ou via un système centralisé. L'idée est d'expérimenter dans cette problématique de suivi agricole les possibilités qu'offre une architecture interopérable. L'application nous est décrite en détail par le CEMAGREF. La discussion essaye naturellement de bien cerner le contexte applicatif. Un contexte P2P a été évoqué dans le cadre des applications avec les agriculteurs.

17h00 Bilan de la journée

- La nécessité de se détacher un peu de l'application, pour enrichir le cadre théorique, est actée. Le CEMAGREF est d'accord, et espère à l'inverse déduire des applications intéressantes de nos contributions.
- Il est décidé que le nombre d'exposés doit être plus limité dans une journée. Des rencontres sur deux jours seraient bien plus propices au travail et aux échanges.
- Zohra propose de définir des sous-groupes de travail par WorkPackage. Pour Farouk, chaque sous-groupe pourrait exposer quelques directions le 16 juin. Mathieu propose que chaque sous-groupe détermine, en particulier, les entrées et les sorties attendues, afin d'assurer une cohérence dans les travaux.
- La date de la prochaine réunion est fixée au 16 juin 2006. Elle aura lieu à Rennes avec pour ordre du jour :
 - o Présentation de Zohra *classification des architectures d'intégration et leurs impacts* en vue de fixer le cadre de l'étude : contexte dynamique ou non ; quel type de requête, etc.
 - o Une présentation par workpackage :
 - o Synthèse des réflexions menées au sein du WP2 et du WP3 (*Chaque coordinateur anime son groupe en vue de la prochaine réunion*)