

Compte-Rendu de la réunion FORUM du 15/05/08

La réunion rassemblait: Fabien Duchateau, Salima Benbernou, Farouk Toumani, Zohra Bellahsene, Olivier Pivert, Hélène Jaudoin, Pierre Colomb, Yannick Loiseau, Samir, François Pinet et Stéphane Bernard.

Invité du Jour : Jérôme Euzénat, INRIA Rhone Apes.

Ordre du Jour:

- Démonstrations par WPi et bilan
- Panification de la prochaine réunion
- Présentation de J. Euzénat, Invité Forum: Alignement d'ontologies
- Présentation de O. Pivert : Présentation de l'article "Pay as you go" de Halevy et al.

1. Démonstrations par WPi et bilan

WP2: Découverte de correspondances

La démonstration est effectuée par Fabien Duchateau. Le prototype prend en entrée des schémas XML et retourne des paires d'attributs pour lesquels un matching a été découvert.

L'essai est réalisé avec 283 schémas comportant environ 30 attributs. Le prototype a retourné plus de 160000 paires d'attributs.

Ce prototype génère également des clusters d'attributs qu'on peut exploiter pour générer un schéma intégré.

Les validations avec le Cemagref sont en cours.

Fabien nous fait part des problèmes relevés lors des tests qu'il a effectué. Le prototype fonctionne mal dès lors que les labels des attributs sont petits. Les issues possibles sont d'utiliser de la connaissance supplémentaire sur le domaine, de prendre en compte des règles linguistiques (notamment ne pas tenir compte des caractères de type "_"), de s'aider du type des attributs ou encore de leur domaine de valeurs (cf les travaux de Manniani).

Fabien souhaite également étudier le problème du passage des matchings aux mappings.

WP4: Évolution de schémas (LIRIS)

Samir nous présente son prototype qui prend en entrée un schéma à l'instant t et ce même schéma à l'instant t+1 et qui décèle les changements opérés sur ce schéma.

Le prototype détecte les éléments qui ont été renommés, supprimés ou déplacés. La prochaine étape à réaliser est compte tenu des mappings existants entre un schéma T et un schéma S à l'instant t, de générer les mappings entre T et S à t+1.

WP3: Réécriture tolérante

Ce prototype prend en entrée les descriptions des vues et une requête en présence de contrainte d'intervalle puis donne en sortie des réécritures dans l'ordre décroissant de leur score qui spécifie pour chaque réécriture R la probabilité que les tuples issus de R soient des réponses certaines de la requête initiale. Il est possible de ne générer que les réécritures qui dépassent un certain score fixé par l'utilisateur.

Pour le moment, le prototype fonctionne dans l'hypothèse de l'indépendance des attributs. La prochaine étape est de mettre en place une solution générale.

Les validations avec le Cemagref sont en cours. Le seul problème qui pourrait se poser est celui du passage à l'échelle.

Bilan général de l'implémentation

On doit rendre le logiciel au Cemagref qui s'occupe de la livraison de la version finale.

Pour cela, le Cemagref doit fournir les vues au WP3 et prévoit une visite au LIRIS. Pour ce qui est du WP2, les parties sont déjà en interaction.

Pour septembre 2008, nous devons écrire un rapport présentant les expérimentations sur tous nos prototypes qui prendront en entrée une quarantaine de sources. Ce rapport se basera sur un scénario que va nous proposer François Pinet.

Concernant la réalisation d'un prototype rassemblant les différents prototypes des WPi, ce projet semble ambitieux au vu du temps qui nous reste.

2. Planification de la prochaine réunion

La prochaine réunion aura lieu le vendredi 27 juin à Montpellier. Les présentations suivantes sont envisagées:

Fabien DeMarchi: Normalisation des sources?

Pierre Colomb: Privacy et intégration

Hélène Jaudoin & Fabien Duchateau: Présentation d'articles sur la découverte de mappings.

François Pinet: Présentation sur les expérimentations en cours

Olivier Pivert: Présentation de l'article de Halevy une deuxième fois?

Réflexions sur l'évaluation des projets.

A l'ordre du jour également, François propose de discuter de la réalisation d'un article commun pour une revue de SIG. Zohra rappelle notre projet de soumettre un article à SIGMOD records sur les thématiques de FORUM. Nous devons aussi nous entretenir sur l'évaluation du projet et demander une prolongation pour la remise du rapport finale.

D'un point de vue livrable, chaque responsable de WPi doit fournir à Zohra, d'ici le 10/06, le rapport d'activité semestriel qui comprend entre autre la liste des publications du premier semestre, les participations au congrès, etc.

La réunion suivante se déroulera pendant les journées BDA 2008 du 22 au 24 octobre.

3. Présentation de J. Euzenat

La présentation de J. Euzenat porte sur le problème de l'alignement d'ontologies. Après avoir présenté l'état de l'art sur la question, il a exposé l'approche suivie dans son équipe.

4. Présentation de O. Pivert

La présentation d'Olivier porte sur la présentation de deux articles d'Halevy et notamment sur une nouvelle architecture d'intégration adaptée au web et sur une approche probabiliste pour répondre aux requêtes par mots clés exprimées dans ce type d'architecture.