

Proposition de sujet de Master M2 2016-2017

Laboratoire : LIAS – ISAE ENSMA / UP

Titre : Découverte et exploitation de dépendances des données de déplacement des populations

Encadrant(s) : Brice CHARDIN*, appuyé par Mickael BARON*, Allel HADJALI* et Cyril ROUSSEL** (*LIAS, **MIGRINTER)

Mots clés : bases de données relationnelles, fouille de données, dépendances fonctionnelles, implications

Pré-requis : programmation orientée objet, ingénierie des données et des modèles, applications distribuées

Contexte et problématique

Dans le cadre du Contrat de Plan État-Région 2015-2020 "innovations NUMERiques au service des Citoyens" (CPER NUMERIC) et plus spécifiquement dans l'axe valorisation des contenus pour l'e-administration, les laboratoires de recherche du site Poitevin travaillent sur un projet multidisciplinaire autour des bases de données utilisables par les décideurs et les citoyens. Ce projet vise à améliorer la connaissance des chercheurs, des décideurs et des citoyens sur les populations réfugiées et déplacées des zones de conflit (Syrie, Irak) à partir de sources de données multiples et diverses (enquêtes de terrain, statistiques d'agences publiques et d'ONG). Le laboratoire LIAS par l'intermédiaire de l'équipe Ingénierie Des Données (IDD) cherche à fournir des outils d'exploitation et d'exploration de ces données en considérant les dépendances logiques entre les informations comme une représentation synthétique et exploitable par l'analyste.

Objectifs

Le laboratoire dispose d'un outil de traitement de requêtes pour la découverte de dépendances dans les bases de données, basé sur un langage appelé RQL¹. L'objectif de ce stage est de développer de nouvelles fonctionnalités au niveau du langage de requête et du calcul des dépendances.

- Prendre en compte les dépendances approximatives par l'intégration d'indices de pertinence.
- Permettre la comparaison d'ensembles de dépendances, pour confronter plusieurs sources de données ou en étudier l'évolution dans le temps.
- Développer des outils avancés de filtrage, classement et visualisation des résultats.

Organisation

De part l'aspect multidisciplinaire de ce projet, les membres du laboratoire et le stagiaire devront échanger avec les analystes du laboratoire MIGRINTER pour comprendre leurs bases de données et prendre en compte leurs besoins applicatifs, en particulier pour identifier les types de dépendances à extraire. Les développements suivront une méthode agile par des validations régulières et des retours des utilisateurs finaux.

Technologies impliquées

- **Analyse lexicale et syntaxique** : JavaCC
- **Services web et interfaces web** : JAX-RS, JSON et AngularJS

1. <http://rql.insa-lyon.fr>

