

Séminaire

Mardi 20 décembre 2016

Data quality in distributed databases: Conditional Functional dependencies Generation in Distributed Data

Soror Sahri

Maitre de conférences, Université Paris, Descartes

Abstract

Conditional Functional Dependencies (CFD) are an extension of Functional Dependencies (FDs) that capture rules about the data consistency. Existing work on discovering CFDs focused on centralized data. We extend this work to horizontally distributed relations. Given an embedded functional dependency, we generate a pattern tableau that represents a CFD. The original feature of our work is generating CFD pattern tableaux from a distributed relation, without merging all the distributed tuples in a centralized relation. We propose a distributed algorithm based on the concept of pattern summary that minimizes data shipping between the sites of distributed relation.

Programme de la journée

9H - 10H30 :

Thématique et état de la recherche

Exposé

Travaux en cours

10H45 - 12H :

Plateforme utilisée (Outils, démo, ...)

Proposition de sujets et problèmes ouverts

Bio

Soror Sahri is associate professor at Université Paris Descartes (Sorbonne Paris Cité) since September 2010 and is a member of LIPADE (Laboratoire d'Informatique Paris Descartes). She received her PhD in computer science from Paris-Dauphine University, where she worked on scalable and distributed databases. Her PhD work was supported by Microsoft Research. It has been presented at Microsoft Academic Days 2006, and at CIDR 2007 (gang show). Actually, her main research concerns especially data management, data quality and distributed databases. She was involved in many projects (Biocase, eGOV-Bus, SquarePredict project, IDV..).