Polytech network form for PhD Research Grants from the China Scholarship Council

This document describes the PhD subject and supervisor proposed by the French Polytech network of 14 university engineering schools. Please contact the PhD supervisor by email or Skype for further information regarding your application.

Supervisor information	
Family name	LAURENT
First name	Anne
Email	anne.laurent@umontpellier.fr
Web reference	http://www.lirmm.fr/~laurent
Lab name	LIRMM - Montpellier Laboratory of Informatics, Robotics and Microelectronics
Lab web site	http://www.lirmm.fr
Polytech name	Polytech Montpellier
University name	University of Montpellier
Country	France

PhD information	
Title	Cliquez ou appuyez ici pour entrer du texte.
Main topics regards to CSC list (3 topics at maximum)	Information processing of intelligent network Structure of new computer systems

	Understanding models and intelligent systems
Required skills in science and engineering	Computer Science ; Data Science ; Web Applications ; System Programming

Subject description (two pages maximum)

- CONTEXT

Networks underly most of the information being processed, should it be from social networks, interactions in life science, smart grids, smart gaz, etc.
For dealing with such data, models and algorithms have been developed. Some are devoted to high-level usages (e.g., data science) while some other ones allow users to process low-level information. For instance, NoSQL and In-Memory models have recently emerged (Neo4j, GraphX,…).

However, it is still difficult to get both efficient access to network information in terms of performances (time consumption) and qualititive and expressive requests such as analytical processing.

In this thesis, we thus aim at developing a new model for scalable graph processing including both low-level performances and high-level features.

The thesis will be co-supervised by Dr. Arnaud CASTELLTORT, Associate Professor at Polytech Montpellier/LIRMM – University of Montpellier, expert in scalable architectures and graph databases.

- SCIENTIFIC OBJECTIVES AND METHOD

In this thesis, we aim at proposing a novel approach to graph databases for managing large scale raw data and allow users to retrieve high-level information (e.g. patterns) very efficiency. For this purpose, we propse to explore a Rust-based architecture [Blandy et al. 2016].

For this purpose, it will be necessary to explore both scientific literature (inductive databases, data science) and cutting-edge technologies (graph database engines, Spark, Rust programming,…) [Malak et al. 2017].

Inductive databases [Bonchi et al. 2005 ; Romel et al. 2011] describe engines that are both able to provide storage and access (read/write) to data in the same time as machine learning or pattern mining features. It has been intensively studied some years ago, especially for extending query languages. Current NoSQL and distributed database engines integrate pattern matching features but do not allow users to express complex queries where the final pattern is not known « a priori ». First attemps have been proposed in our first works and need to be extended [Castelltort et al. 2017].

For this purpose, temporal data and their representation will be studied as such data is crucial for machine learning [Castelltort et al. 2015].

- REFERENCES

[Blandy et al. 2016] Jim Blandy. Programming Rust. O'Reilly. 2016

[Bonchi et al. 2005] Francesco Bonchi, Jean-François Boulicaut:. Knowledge Discovery in Inductive Databases, 4th International Workshop, KDID 2005, Porto, Portugal, October 3, 2005, Revised Selected and Invited Papers. Lecture Notes in Computer Science 3933, Springer 2006

[Castelltort et al. 2017] Arnaud Castelltort, Anne Laurent. Exploiting NoSQL Graph Databases and in Memory Architectures for Extracting Graph Structural Data Summaries. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 25(1): 81-110 (2017)

[Castelltort et al. 2015] Arnaud Castelltort, Anne Laurent. Fuzzy Historical Graph Pattern Matching A NoSQL Graph Database Approach for Fraud Ring Resolution. AIAI 2015: 151-167

[Malak et al. 2017] Michael S. Malak. Spark GraphX in Action. Manning Publications. 2017

[Romel et al. 2011] Andrea Romei, Franco Turini. Inductive database languages: requirements and examples. Knowledge and Information Systems. Vol. 26. Issue 3. 2011.