

INTERROGATION FLEXIBLE DE BASES DE DONNÉES GRAPHE : ÉTUDE ET MISE EN ŒUVRE

DATES : mars à août 2015, pour une durée approximative de 5 mois.

LIEU : IRISA/SHAMAN, Lannion (Univ. Rennes 1).

ENCADREMENT : Olivier Pivert (pivert@enssat.fr) & Virginie Thion (virginie.thion@irisa.fr).

CONTEXTE

En raison du volume exponentiellement croissant de données ouvertes disponibles, l'intérêt pour les systèmes de gestion de bases de données graphe est grandissant. Ces systèmes, dans la mouvance des systèmes NoSQL (*Not Only SQL*), permettent de manipuler des données structurées sous forme de graphe, auxquelles aucun schéma n'est a priori associé. Ainsi, les systèmes de gestion de bases de données graphe permettent entre autres la gestion de données de réseaux sociaux, d'images, de vidéos, de données bioinformatiques, et de données issues du *semantic web* telles que des données au format RDF éventuellement issues de l'exploitation de données ouvertes. De nombreux systèmes permettant la gestion de bases de données graphe sont apparus. Il s'agit par exemple de AllegroGraph [1], InfiniteGraph [2], Neo4j [3] ou Sparksee [5].

Interroger ce type de données sans schéma associé est une tâche difficile pour un utilisateur, qui se retrouve finalement souvent à poser des requêtes « en aveugle » dont les réponses peuvent être pléthoriques ou vides. Une façon de simplifier la tâche de l'utilisateur dans ce cas est de lui permettre de poser des requêtes « approximatives » telles que :

- « *Trouver les personnes proches de Billy dans le graphe* », où *proche* est un terme approximatif aussi dit *flou* ;
- « *Trouver les personnes qui aiment beaucoup d'autres personnes (ou qui sont aimées par beaucoup d'autres personnes)* », où *beaucoup* est un terme flou.

Le stage se situe dans ce contexte, celui de l'interrogation flexible de bases de données graphe.

OBJECTIFS DU STAGE

Les objectifs du stage sont les suivants :

(Objectif 1) La première indispensable et cruciale tâche consiste en l'étude de la littérature sur le sujet, incluant la revue d'articles de recherche et d'outils existants ;

Vient ensuite l'élaboration de la contribution scientifique du stage avec :

(Objectif 2) une facette théorique consistant en la définition ou l'enrichissement d'un langage de requêtes flexibles permettant d'interroger des bases de données graphes ;

(Objectif 3) une facette plus appliquée consistant en l'implantation d'un prototype, fondé sur un système de gestion de données existant à choisir, permettant la mise en œuvre du langage de requête précédemment défini.

Le sujet du stage pourra éventuellement être amené à évoluer en fonction des outils utilisés et des centres d'intérêt manifestés par l'étudiant.

BÉNÉFICES POUR L'ÉTUDIANT

Ce stage permettra :

- une montée en compétences sur les outils de bases de données graphe de plus en plus étudiés dans le monde de la recherche, et de plus en plus utilisés en entreprise.
- la découverte du monde de la recherche, en étant intégré à une équipe de recherche le temps du stage (équipe Shaman de l'IRISA [4] localisée dans les locaux de l'ENSSAT à Lannion sur la superbe côte de Granit Rose).

ENCADRANT/CONTACT

Les encadrants du stage sont Olivier Pivert (pivert@enssat.fr) et Virginie Thion (virginie.thion@irisa.fr). Les contacter pour tout renseignement ou en joignant CV et lettre de motivation pour toute candidature.

RÉFÉRENCES

- [1] AllegroGraph web site. <http://franz.com/agraph/allegrograph>.
- [2] InfiniteGraph web site. <http://www.objectivity.com/infinitegraph>.
- [3] Neo4j web site. <http://www.neo4j.org/>.
- [4] Shaman web site. <http://www-shaman.irisa.fr>.
- [5] Sparksee (formerly known as DEX) web site. <http://sparsity-technologies.com>.