

Sujet de PFE en Master

Titre: Mise en oeuvre de techniques de fouille de données dans une base orientée graphe: application aux systèmes de questions / réponse

Proposé par: D. BOUKRAA

Description : L'objectif de ce travail est de permettre l'extraction de connaissances à partir d'une base de données orientée graphes (exemple: Neo4j). Le travail consiste à modéliser et implémenter les données de questions/ réponses et entraide (ex: StackOverflow) sous la forme d'un graphe de données, puis d'implémenter différents algorithmes tels que la détection de communautés d'utilisateurs, le page-ranking des utilisateur et des tags, etc.

Le travail applique des algorithmes déjà implémentés sous Neo4j pour différents objectifs

Pour tous les utilisateurs

- Détecter les groupes d'utilisateurs selon des critères de ressemblance
- Détecter et classer les utilisateurs selon leur influence dans les réseaux d'entraide en se basant sur le page-ranking
- Détecter les groupes de tags utilisées dans les questions / réponses

Pour un utilisateur donné

- Recommander des questions à l'utilisateur
- Connaître son influence dans sa communauté ou dans l'ensemble du réseau

Le travail nécessite d'abord de représenter les données sous la forme de graphes reliant les utilisateurs et les tags autour de l'élément pivot "posts" et en exploitant les liens de commentaires, des votes, de questions et de réponses.

Le travail sera appliqué au sous-réseau d'entraide informatique "stackoverflow" du réseau stack mais doit pouvoir être applicable aux sous-réseaux similaires (ex: askubuntu, ...)