

STAGE LIFO-LIFAT-LIRIS (groupe de travail DOING)

ACRONYME: stage-DSQ@DOING

Titre : VERS DES REQUÊTES DATA SCIENCE

Disciplines : informatique (bases de données, science de données et intelligence artificielle).

Stage dans le cadre de l'action **DOING (Données Intelligentes)** du **GDR MADICS** et du réseau régional DIAMS (Réseau Thématique de Recherche Données, Intelligence Artificielle, Modélisation et Simulation). Le stage est centré sur une nouvelle collaboration des chercheurs en bases de données et intelligence artificielle du LIFO, du LIFAT et du LIRIS.

Encadrement :

- Mirian Halfeld Ferrari (LIFO, Orléans), mirian@univ-orleans.fr
- Genoveva Vargas-Solar (LIG, Grenoble), genoveva.vargas-solar@univ-lyon1.fr
- Donatello Conte (LIFAT, Tours), donatello.conte@univ-tours.fr
- Jean-Yves Ramel (LIFAT, Tours), jean-yves.ramel@univ-tours.fr

Sujet du stage : DOING s'intéresse à la mise en place des requêtes *data science sur de bases de données graphe*, c'est-à-dire, des requêtes capables de combiner des méthodes d'analyse (par exemple, via des algorithmes d'apprentissage) avec des méthodes standards (par exemple en regroupant des informations à afficher à l'utilisateur). Le stage a comme objectif l'étude de la conception d'un système d'interrogation capable d'offrir à ses utilisateurs des requêtes d'analyse prescriptive (*Prescriptive Analytics* en anglais). Ce terme désigne la nouvelle frontière de l'analyse de données, qui combine la gestion des données (dans le sens des bases de données), le *machine learning*, et la recherche opérationnelle.

Nous sommes particulièrement intéressés par une approche liée au langage Cypher, sur des graphes de propriétés. Dans ce cadre, l'extraction des données se fait par la recherche de nœuds du graphe qui satisfont certaines caractéristiques sur les attributs et sur les relations avec les autres nœuds. Au cœur de ce modèle de prédiction il y a donc les méthodes de classification des nœuds d'un graphe. Nous voulons étudier le problème de la classification de nœuds par des techniques d'apprentissage profond, et proposer une adaptation de ces méthodes pour qu'il soit possible son intégration dans un système de gestion de bases de données orientée graphes et utilisant un langage Cypher-like.

Le travail concerne principalement les points ci-dessous:

1. **Faire un état de l'art englobant:**
 - des méthodes de classification de nœuds par apprentissage profond;
 - des propositions pour intégrer requêtes et algorithmes d'apprentissage;
 - les possibilités d'analyse du langage Cypher et sa bibliothèque *data science*.
2. **Modéliser une requête *Cypher like qui englobe* une méthode de classification de nœuds sur graphe**
3. **Adapter une méthode de classification de nœuds au contexte défini au point 2**

Lieu du stage: LIFO (Orléans, Mirian Halfeld-Ferrari encadrant référent), avec des déplacements possibles au LIFAT (Tours). Nous souhaitons que le stagiaire soit physiquement au LIFO, même si des périodes de télétravail peuvent être envisagés/ programmés.

Profil souhaité: l'étudiant doit être intéressé par l'approfondissement de ses connaissances dans le domaine de bases de données et intelligence artificielle. Niveau **master**.

Durée du stage: de **4 à 6 mois**, financé par les fond d'un contrat CIFRE Orléans.

CANDIDATURE:

Etape 1: Les candidats doivent envoyer (1) Curriculum Vitae et (2) Relevé de notes aux encadrants (emails ci-dessus) avant le **15 DECEMBRE 2021 avec l'acronyme du stage en sujet du email**.

Etape 2: Candidats sélectionnés seront contactés (vers le 16 décembre) pour fixer pour début janvier un entretien par visio (l'entretien inclura une petite discussion sur un article, donné en amont).