

Stage de Master au laboratoire ERIC

Recommandation de liens dans un réseau d'auteurs

Contact : julien.velcin@univ-lyon2.fr

Durée : 4 à 6 mois

Gratification : environ 600 euros par mois

Localisation : Laboratoire ERIC, Université Lyon 2 (campus de Bron)

Début du stage : Mars-Avril 2020 (possibilités de commencer plus tôt si besoin)

Mots clefs : data science, natural language processing, machine learning, representation learning, topic modeling, digital humanities

1 Contexte général

Le projet LIFRANUM (Littératures FRAncophones Numériques), financé par l'Agence Nationale de la Recherche pour une période allant de 2020 à 2023, vise à analyser l'impact des modifications des supports de l'écrit sur les pratiques littéraires. Il regroupe des chercheurs de l'équipe MARGE (Université Lyon 3) et du laboratoire ERIC (Universités Lyon 2 et Lyon 1), en partenariat avec la Bibliothèque nationale de France. L'ambition du projet est notamment de construire une plateforme en ligne qui permettra de faciliter l'analyse des créations littéraires sur le Web par les chercheurs en information-communication et en littérature.

2 Objectif du stage

Dans le cadre du projet LIFRANUM, nous cherchons un stagiaire en Informatique afin de pouvoir tester plusieurs algorithmes de la littérature (informatique) pour suggérer de nouveaux liens dans le réseau des auteurs des textes produits sur les supports numériques (par ex. site web personnel, blog), mais également pour proposer des algorithmes originaux. Pour bien débuter, le stagiaire pourra s'appuyer sur des travaux en cours menés par des chercheurs du laboratoire ERIC sur l'apprentissage de représentation (*representation learning*) d'auteurs. Il aura accès à des données récupérées dans le cadre du projet, mais il travaillera également sur un jeu de données déjà collectées au sein du laboratoire.

3 Ce qu'il faut faire

Les tâches prioritaire à réaliser sont les suivantes :

- Etudier les travaux récents en apprentissage de représentations à partir de réseaux de documents textuels, en particulier ceux développés au sein du laboratoire (partie état de l'art).
- Prendre connaissance des données fournies pour le stage (banc d'essai et données du projet).
- Mettre en pratique des méthodes basées sur l'apprentissage de représentation et sur la modélisation thématique afin de suggérer des liens entre documents et auteurs.
- Implémenter au moins une méthode originale développée par les chercheurs du laboratoire, en collaboration avec le stagiaire.
- Intégrer cette solution dans un logiciel actuellement en cours de développement au sein du laboratoire.

4 Profil recherché

Nous recherchons un candidat ayant des compétences solides en analyse/fouille de données, en programmation (Python de préférence) et si possible des notions de traitement automatique des langues (*natural language processing*) / fouille de données textuelles (*text mining*) et d'apprentissage automatique (*machine learning*). Un intérêt pour le travail pluridisciplinaire serait un plus.

5 Déroulement du stage

Le stage se déroulera dans les locaux du laboratoire ERIC, sur le campus de Bron (à environ 30 min. du centre-ville de Lyon en tram). Il sera encadré par un enseignant-chercheur permanent accompagné par un étudiant en Doctorat travaillant sur les thématiques du stage. Une réunion hebdomadaire est prévue, en plus des réunions organisées pour le projet LIFRANUM.

6 Poursuite du stage possible

Un financement de thèse pour 3 ans est prévu à partir de septembre 2020. Si le stage donne satisfaction et que l'étudiant est intéressé, celui-ci constitue un candidat privilégié pour obtenir cette allocation.

7 Procédure de candidature

Les candidats doivent envoyer les documents suivants à l'adresse julien.velcin@univ-lyon2.fr avant le 15 janvier 2020 :

- CV
- lettre de motivation
- relevés de notes des deux dernières années

Les candidats retenus seront convoqués pour un entretien durant la deuxième quinzaine du mois de janvier. Les résultats devraient être communiqués dans la première semaine de février.