



Stage de Master

[Projet ANR LIFRANUM](#)

Conception et mise en œuvre de l'interface d'analyse d'un lac de données web

Lieu du stage : [Laboratoire ERIC](#), Université Lyon 2 (campus Porte des Alpes, Bron)

Encadrement : Jérôme Darmont et Sabine Loudcher

Durée : 5 à 6 mois à partir de février 2022

Rémunération : Indemnités de stage légales (environ 550 € par mois)

Contexte : Le projet LIFRANUM (Littératures FRAncophones NUMériques), porté par le [laboratoire MARGE](#), vise à identifier, indexer et analyser des productions littéraires nativement numériques dans l'aire francophone. Pour cela, l'outil de référence de l'archivage web Heritrix a en premier lieu permis de constituer un corpus sous forme de fichiers de conservation au format Web ARChive (WARC). Des métadonnées des pages HTML ont ensuite été extraites des fichiers WARC (contenu textuel, fichier PDF, images, vidéo, etc.) et indexées dans Solr.

Par ailleurs, en s'appuyant sur des blogs appartenant à des auteurs identifiés, un deuxième corpus (fichiers JSON) a été constitué via les API de Wordpress et Blogger. Des informations, pages, posts et commentaires ont été extraits et constituent la série de métadonnées, métadonnées stockées et indexées dans MongoDB.

Sujet et missions : L'objectif de ce stage est de concevoir et mettre en œuvre une interface web commune aux deux types de sources de métadonnées (celles issues des WARC et celles provenant des API de blogs) pour permettre aux chercheur·es du laboratoire MARGE de requêter et d'analyser les données sous-jacentes. Il faudra pour cela :

- étudier l'architecture de données déjà en place ;
- concevoir un schéma d'alignement des métadonnées des WARC et des API ;
- concevoir et mettre en œuvre une interface graphique permettant de rechercher, via les métadonnées consolidées, les données (à la manière d'un moteur de recherche) ;
- proposer des visualisations « toutes faites » (mais paramétrables) ou ad-hoc, en lien avec les chercheur·es du laboratoire MARGE.

Compétences souhaitées : technologies *big data*, programmation web, gestion de données, *data visualization*, rigueur.

Pour candidater : Envoyer un CV, une lettre de motivation et vos notes des années universitaires précédentes à jerome.darmont@univ-lyon2.fr et sabine.loudcher@univ-lyon2.fr avec en objet « Candidature stage LIFRANUM » d'ici le **15 novembre 2021**.