

Type de recrutement : **Ingénieur de recherche**

Niveau d'emploi : IGR

Localisation du poste : Lyon

Contexte et environnement de travail

Contexte local

Le poste s'inscrit dans le cadre du projet ANR LIFRANUM 2020-2023 (Littératures Francophones Numériques). Le projet LIFRANUM veut identifier et structurer le corpus des littératures francophones nativement numériques (sites, blogs, réseaux sociaux). Cette dimension patrimoniale se double d'une interrogation épistémologique sur la littérarité des contenus repérés et la dynamique des sociabilités identifiées. Le projet s'appuie sur deux laboratoires englobant littérature, info-com et informatique ; il acte la collaboration de la BnF ; il bénéficie du soutien de l'Institut International de la Francophonie. L'objectif du projet concerne donc la communauté littéraire, mais bien au-delà, vise à rendre disponibles à tous les champs disciplinaires un corpus d'ampleur ainsi qu'une méthodologie novatrice.

Objectifs du recrutement

Il s'agit pour cet ingénieur d'identifier, de répertorier et de crawler les contenus des sites web, blogs, réseaux sociaux francophones proposant un contenu littéraire nativement numérique. L'ingénieur aura à identifier ces productions (quelle que soit la plateforme de diffusion – et notamment Facebook) et à les répertorier avant le crawl. Il aura à contribuer à définir la stratégie de crawling. Les résultats de ces crawls seront stockés dans une base de données et indexés en métadonnées WARC. Une part essentielle du travail consistera alors à enrichir sémantiquement ces métadonnées de façon à faciliter recherche et navigation dans le corpus.

Positionnement et bénéfices attendus

Le travail de l'ingénieur est à la fois *technique* au sens où il requiert une maîtrise des formats et des templates du web, les outils de crawling et l'archivage web, et il est également *conceptuel* dans la mesure où il faudra caractériser et catégoriser les pages crawlées. En ce sens, l'enrichissement sémantique constitue une part importante du travail.

L'ingénieur aura ainsi à amorcer l'élaboration du corpus, au sens où les choix qui seront faits contraindront la suite du développement du projet. Comme il s'agit de constituer un corpus, nous devons respecter les règles de complétude et d'exhaustivité (au minimum de représentativité) associées à cet exercice initié par les linguistes.

Mission du poste et activités principales

L'ingénieur aura comme missions :

- Fouiller patiemment le web de façon à identifier ces contenus (et tous les liens qui leur sont associés), avant de procéder au crawling.
- Identifier, répertorier et crawler les pages et sites contenant un contenu littéraire nativement numérique selon une stratégie à définir.

- Enrichir les métadonnées associées au crawl de façon à faciliter la recherche et la navigation.
- Contribuer à construire le langage de description et d'échange des objets.

Profil recherché

Doctorat ou diplôme d'ingénieur, avec spécialisation dans les technologies de l'information et web.

Une connaissance approfondie du web sémantique, de l'archivage électronique et de la description documentaire (langages documentaires, terminologies).

Des connaissances en « Humanités Numériques » et en littérature sont nécessaires.

Compétences recherchées

Maîtrise des techniques du crawling et des systèmes d'exploitation LINUX et WINDOWS.

Bonne maîtrise de l'anglais.

Maîtrise des langages de programmation Python et de développement JAVA et des langages du web RDF, OWL, SKOS.

Maîtrise des jeux de métadonnées et des ontologies bibliographiques.

Connaissance minimale des bases de données de type « lac de données ».

⇒ *Savoir :*

Des connaissances en littérature sont requises, de même qu'un intérêt marqué pour les SHS. Les problématiques des « humanités numériques », celles de la construction de corpus à partir du web, celles de la représentation et de l'organisation des données, doivent être connues.

Rigueur dans la réalisation des tâches : cela concerne à la fois la dimension corpus et la dimension identification de l'information.

La connaissance de jeux de données dans le domaine des SHS, et si possible en littérature, serait un plus.

⇒ *Savoir-faire*

Haut niveau de technicité requis.

Conduite de projet : l'ingénieur doit être capable de se projeter sur toutes les phases de développement du projet, de façon à prévoir dans son activité présente les implications pour la suite du projet.

Rédaction : capacité à exprimer et expliquer les idées à la fois à l'oral et à l'écrit (y compris en anglais).

Capacité réflexive importante, notamment relativement à ses propres méthodes et méthodologies.

Capacité à travailler de façon inductive.

Adaptabilité : l'ingénieur aura à traiter à la fois du matériau littéraire, des communautés d'auteurs identifiables sur le web, des formats et templates hétérogènes, des outils de description et des techniques de crawling. Il devra donc faire preuve d'une forte capacité d'adaptation et de compréhension d'univers scientifiques différents.

Patience : l'identification des données est un travail long et dont les résultats ne sont pas systématiquement visibles immédiatement. Remise en question et tâtonnement font partie du quotidien de ce travail.

Date de prise de poste : 1^{er} février 2020. Durée : 12 mois

Localisation :

Équipe MARGE, université Lyon 3. L'ingénieur disposera d'un bureau dédié à la MSH de Lyon-Saint Étienne, 14 av. Berthelot, 69007 LYON.

Rémunération : grille IGR

Merci d'envoyer CV détaillé et lettre de motivation par courriel aux adresses suivantes : Frederique.lozanorios@univ-lyon3.fr, gilles.bonnet@univ-lyon3.fr, christian.cote@univ-lyon3.fr pour le 31 décembre 2019 délai de rigueur. Les entretiens auront lieu à Lyon les 9 et 10 janvier 2020, pour une prise de poste le 1^{er} février 2020.