CDD 5 mois - Ingénieur d'étude / de recherche – Fouille de données pour qualification de données historiques

Missions

AU CNRS, le laboratoire Littoral Environnement et Société (Unité Mixte de Recherche n° 7266) à la Rochelle propose un poste d'**Ingénieur d'Etude ou de Recherche en informatique** avec 0 à 5 années d'expérience. La mission consiste à implémenter et tester diverses méthodes de contrôles de cohérence sur une base de données historiques concernant le commerce maritime du XVIIIème siècle. Il s'agit de croiser deux corpus de données historiques imparfaites concernant le commerce maritime du XVIII ème siècle et la balance du commerce. Les humanités numériques accompagnent toutes les étapes du projet ANR PORTIC, contexte de cette mission. Les approches numériques qui seront implémentées ont pour objectif de tagger le corpus afin de distinguer différents niveaux de qualité dans les données. La mission qui commence le 15 Avril 2019 pour une durée de 5 mois se fera en interaction aussi avec Alain Bouju, maître de conférences en informatique du L3I, à la Rochelle.

Activités

L'ingénieur(e) réalisera sa mission de façon créative et autonome dans un milieu fortement interdisciplinaire, avec des outils open-source. Il s'agit de développer des services de qualification de données au dessus d'une base de données ouvertes. L'ingénieur(e) :

- Participera à l'élaboration d'outils de consultation, de contrôle et de gestion (scripts, procédures, requêtes, reporting)

- Assistera la maîtrise d'ouvrage dans l'élaboration de cahiers des charges

- Rédigera la documentation fonctionnelle et technique

L'ingénieur utilisera des méthodes de fouille de données pour l'identification d'aberrations ou d'incohérence dans le corpus. Il s'agit ici de réutiliser des bibliothèques open-source: méthodes de classification non supervisées pour la détection de patrons et de similarités, statistiques non paramétriques pour la détection de valeurs exceptionnelles. Concrètement, en concevant et programmant une application permettant d'activer ces méthodes et d'exploiter leurs sorties pour tagger les données, la contribution de l'ingénieur permettra de connaître finement le niveau de qualité des données.

Enfin, suivant les résultats obtenus, le travail pourra faire l'objet de publications scientifiques et l'ingénieur(e) sera associé(e) étroitement à cette tâche, comme co-auteur(e) ou auteur(e) principal(e) le cas échéant. Ce poste offre donc l'opportunité d'apprendre à rédiger une communication scientifique, sur différents supports (article, poster) et permettra de valoriser les travaux réalisés pour l'avenir puisque le code sera diffusé sous licence libre.

Compétences attendues

Le candidat doit être diplômé avec un master en informatique ou comme ingénieur en informatique, ou en sciences de l'information.

- Le poste nécessite des compétences confirmées en programmation (Service REST, Java, Python) ainsi qu'en base de données (Postgres, RDF, graphes, UML).

- Des compétences seraient appréciées en mathématiques et statistiques, et dans l'une ou plusieurs de ces technologies : Strasbon, Postgis, Neo4J, R.

- Le poste requiert un goût pour le dialogue avec les différents acteurs du projet et leurs diverses disciplines (histoire, économie, géomatique, informatique).

Contexte de travail

L'ingénieur(e) intégrera l'Unité Mixte de Recherche Littoral Environnement et Sociétés, (U.M.R. 7266 LIENSs), 200 personnes. Ce laboratoire regroupe des experts scientifiques des disciplines de l'écologie, la géographie, la biologie, l'histoire, la chimie moléculaire et les sciences de la terre et interroge des questions liées au développement durable et au changement climatique autour des zones littorales (<https://lienss.univ-larochelle.fr/>). L'ingénieur(e) intégrera donc un milieu fortement interdisciplinaire et en particulier le service DISA, une plateforme base de données sous la responsabilité scientifique de Christine Plumejeaud-Perreau. DISA est au croisement de nombreux projets scientifiques afin d'offrir une meilleure capacité de croisement de données fortement hétérogènes et de favoriser la mise en œuvre des principes FAIR dans la recherche.

L'équipe de PORTIC sur la Rochelle est coordonnée par Christine Plumejeaud-Perreau, informaticienne responsable de la plateforme DISA au LIENSs. Elle travaille depuis 5 ans avec Alain Bouju, Maître de conférences avec Habilitation à Diriger des Recherches en informatique au **Laboratoire d'Informatique, Images et Interactions (L3i)** de l'Université de la Rochelle depuis 2014. L'ingénieur(e) sera donc en interaction forte avec ces deux personnes pour mener sa mission.

Il ou elle disposera d'un portable pour le développement et accèdera à toutes les facilités qu'offre la Très Grande Infrastructure de Recherche HumaNum (machine virtuelle pour déploiement en production des outils développés, GIT, espace de stockage). Tout ce qui sera développé par PORTIC sera sous licence libre. Ce contrat à durée déterminée est envisagé comme un prélude à un doctorat, dont l'offre est en ligne également.

Contraintes et risques

Sans

Informations contractuelles

- Durée du contrat : 5 mois
- Quotité : temps complet
- Date de publication: 17 décembre
- Date d'embauche: 1er avril 2019
- Diplôme souhaité : Bac+5
- Expérience : entre 0 et 5 ans.
- Rémunération indicative : entre 1984 et 2310 euros bruts mensuels selon qualification.

La candidature doit inclure un CV, une lettre de motivation, un relevé de notes de Master 2 ou école d'ingénieur dernière année, ainsi que le dernier rapport de stage. Postuler en ligne sur :

<https://emploi.cnrs.fr/Offres/CDD/UMR7266-CHRPLU-001/Default.aspx>

Short-term contract 5 months – Engineer – Datamining for historical data qualification

Missions

In the Littoral Environment and Society laboratory (Mixed Research Unit n ° 7266) at La Rochelle, the CNRS proposes a position of Engineer of Study or Research in computer science with 0 to 5 years of experience. The engineer will have to implement and test various methods of consistency checks on a historical database of eighteenth-century maritime commerce, in the context of the PORTIC project, funded by the French National Agency for Research (ANR). It is a question of crossing two corpus of imperfect historical data concerning the eighteenth century maritime trade and the balance of trade. Digital humanities accompany all stages of the project. The numerical approaches that will be used are aimed at tagging the corpus in order to distinguish different levels of quality within the data. The mission begins on April 15, 2019 for a period of 5 months, and will be realized also in interaction with the L3I computer science laboratory, at La Rochelle.

Activities

The engineer will perform his mission in a creative and autonomous way in a highly interdisciplinary environment, with open-source tools. It involves developing data qualification services over an open database. The engineer will use data mining methods to identify aberrations or inconsistency in the corpus. This involves reusing open-source libraries: unsupervised classification methods for pattern and similarity detection, non-parametric statistics for the detection of exceptional values. Specifically, by designing and programming an application to activate these methods and exploit their outputs to tag the data, the contribution of the engineer will help to determine finely the various levels of data quality inside the corpus.

Finally, depending on the results obtained, the work may be the subject of scientific publications and the engineer will be closely associated with this task, as co-author or principal author, where appropriate. This post therefore offers the opportunity to learn to write a scientific paper, on different media (article, poster) and will enhance the work done for the future since the code will be released under free license.

Candidate profile

Candidate should have a master or an engineer diploma in computer sciences.

- The position therefore requires programming skills (REST Service, Java, Python) as well as database (Postgres, RDF, graphs, UML).
- Skills would be appreciated in mathematics/statistics and one or more of these technologies: Strasbon, Postgis, Neo4J, R.
- The position requires a continuous dialogue with the various actors of the project and their various disciplines (history, economics, geomatics, computer science).

Work context

The engineer will integrate the Mixed Research Unit Littoral Environment and Societies, (U.M.R. 7266 LINKS), 200 persons. This laboratory brings together scientific experts from the disciplines of ecology, geography, biology, history, molecular chemistry and earth sciences and questions issues related to sustainable development and climate change around coastal areas ([https : //lienss.univ-larochelle.fr/](https://lienss.univ-larochelle.fr/)). The engineer will thus integrate a highly interdisciplinary environment and in particular the DISA service, a database platform at the crossroads of numerous scientific projects in order to offer a better capacity for cross-referencing highly heterogeneous data and to promote the implementation of FAIR principles in research.

The PORTIC team in La Rochelle is coordinated by Christine Plumejeaud-Perreau, computer scientist responsible for the DISA platform at LIENSs. She has been working for 5 years with Alain Bouju, Associate Professor with Habilitation to lead Research in Computer Science at the Laboratory of Computer Science, Images and Interactions (L3i) at the University of La Rochelle since 2014. The engineer will therefore be in strong interaction with these two people to carry out his mission.

He or she will have a laptop for development and will access all the facilities offered by the Very Large Research Infrastructure HumaNum (virtual machine for deployment in production tools developed, GIT, storage space). All its work will be released by PORTIC under free license. This fixed-term contract is envisaged as a prelude to a PhD, which is also online.

Constraints and risks

None

Contractual information

- Duration of the contract: 5 months
- Quotity: full time
- Publication date: December 17
- Date of hiring: April 1st, 2019
- Desired degree: Bac + 5 in computer sciences or mathematics
- Experience: between 0 and 5 years old.
- Indicative remuneration: between 1 984 and 2 310 euros gross monthly according to qualification

The application must include a CV, a letter of motivation, a transcript of Master 2 or engineering school last year, as well as the latest internship report. Apply online : <https://emploi.cnrs.fr/Offres/CDD/UMR7266-CHRPLU-001/Default.aspx?lang=EN>