



L'Université Paul-Valéry Montpellier 3 recrute un-e Chargé.e d'analyse de corpus et de données avec R

Catégorie : A, **par référence au niveau IGE**

CDD de 12 mois

Quotité : **100 %**

Prise de fonction : 01/01/2022

L'Université Paul-Valéry Montpellier 3 est une université des Humanités : Lettres, Langues, Arts, Sciences Humaines et Sociales. Plus de 20 000 étudiants sont inscrits de la licence au doctorat et 240 diplômés en lien avec les besoins socio-économiques sont préparés.

Contexte

Numapresse est un projet ANR qui ambitionne de proposer une nouvelle histoire culturelle et littéraire de la presse à partir de l'application de nouvelles méthodes d'analyses automatisées de fouilles de données à de grands corpus patrimoniaux numérisés depuis le 19e siècle jusqu'à aujourd'hui. Depuis 2017, le projet a développé des outils inédits pour identifier les genres journalistiques, les illustrations de presse, les signatures de journalistes. Ces méthodes ont rendu possible l'exploration de corpus difficiles d'accès comme la page de cinéma dans les quotidiens de l'entre-deux-guerres ou les magazines des années 1950.

Les collections et les outils de Numapresse sont principalement hébergés sur un espace R Studio mis à disposition par Humanum. Le projet a été amené à collecter un grand nombre de documents numérisés mis à disposition par des bibliothèques patrimoniales. Il a également lancé des chantiers de numérisation internes de corpus imprimés anciens.

L'équipe qui porte le projet souhaite recruter un collaborateur.trice afin d'assurer la continuité des opérations techniques et scientifiques.

Missions et activités

Dans le cadre de ses fonctions, il/elle assure les missions suivantes :

- exécution et entretien de scripts d'analyses de corpus patrimoniaux en R et Python.
- importation de corpus de presse et de périodiques mis à disposition par des bibliothèques numériques (Gallica, Ressources Patrimoines Occitanie).
- organisation de la numérisation de collections internes de presse et de périodiques, en association avec des vacataires étudiant.es : création des scans, OCRisation, intégration des métadonnées.
- publication de corpus et des données traités sur des interfaces en PHP/MySQL.
- participation à l'organisation des séances d'atelier et de séminaire et des conférences du projet.
- rédaction de la documentation du projet et de notes de recherche sur le carnet du projet
- relation entre les équipes, les fournisseurs de services (Huma-Num)

Compétences requises

Le candidat ou la candidate dispose des connaissances, savoir-faire et aptitudes suivants:

- maîtrise de R, du tidyverse et de l'interface RStudio.
- usage au moins occasionnel de Python.
- connaissance des formats patrimoniaux (XML, METS, IIIF) et des logiciels de développement web dans le domaine patrimonial (PHP/MySQL, Omeka, Wordpress)
- familiarité avec les programmes d'importation automatisée de fichiers (Rsync)
- familiarité avec les outils de numérisation et d'OCR (tesseract)

Les compétences suivantes sont également souhaitées mais non requises :

- expérience dans la conduite ou l'exécution de projets en humanités numériques
- familiarité avec les grandes bases de données patrimoniales françaises (Data BNF, API BNF...) et plus largement avec les sciences de l'information et des bibliothèques
- initiation au machine learning appliqué à des corpus textuels ou visuels.
- notions d'histoire de la presse du XIXe et du XXe siècle
- capacité à vulgariser des opérations informatiques auprès d'un public non initié
- autonomie, compétences relationnelles;

Formations et/ou expérience professionnelle souhaitables :

Master 2 dans les domaines de l'ingénierie informatique, de l'ingénierie numérique ou des humanités numériques ou d'un Master 2 en sciences humaines et sociales avec une formation renforcée en analyse de données (R, Python)

Lieu d'exercice : Au sein de l'équipe du projet Numapresse, sous la responsabilité de Marie-Ève Thérenty, le candidat ou la candidate est chargé.e du développement des outils, du suivi des séminaires, du dialogue avec les équipes du projet.

Le poste sera principalement localisé à Montpellier (avec la possibilité de télétravail deux jours par semaine) sous la responsabilité de Julien Schuh et Marie-Eve Thérenty)

Rémunération : entre 1840 euros et 2295 euros brut mensuel selon l'expérience acquise dans des fonctions équivalentes.

Les candidatures sont à adresser selon les modalités suivantes :

Le dossier de candidature complet devra être adressé, sous format numérique en un seul PDF, au plus tard

le 27 novembre 2020 à minuit à l'adresse suivante: marie-eve.therenty@univ-montp3.fr.

Les candidats seront auditionnés en visioconférence durant le mois de décembre.

Le dossier de candidature comprend :

- un curriculum vitae et une liste de travaux/publications
- le diplôme de master 2
- le rapport de thèse (le cas échéant)
- une lettre de motivation
- optionnellement, une liste de réalisations antérieures en analyse de données/humanités numériques.