Maison des sciences humaines et sociales en Bretagne

Recrutement d'un ingénieur ou d'une ingénieure en traitement et analyse de corpus textuels

La Maison des sciences humaines et sociales en Bretagne recrute par voie contractuelle un ingénieur ou une ingénieure en traitement et analyse de corpus textuels pour la plateforme Humanités Numériques.

Contexte de travail

La Maison des Sciences humaines et sociales en Bretagne est une Unité d'appui et de recherche (UAR 3549) qui accompagne les unités bretonnes de recherche en arts, lettres, langues, sciences humaines et sociales.
Elle est membre du Réseau national des Maisons des Sciences sociales et des Humanités (RnMSH) et compte six tutelles : le CNRS, l'EHESP, l'université de Bretagne occidentale, l'université de Bretagne Sud, l'université de Rennes et l'université Rennes 2.
Au sein de la plateforme Humanités numériques, l'ingénieur ou l'ingénieure en traitement et analyse de corpus textuels travaillera, sous la responsabilité hiérarchique de la responsable de la plateforme Humanités numériques.

Missions

L'ingénieur l'ingénieure en traitement et analyse de corpus textuels a pour mission d'accompagner les projets de recherche soutenus par la Maison des Sciences humaines et sociales en Bretagne dans la production, le traitement et la diffusion des données de la recherche en SHS.
Il sera amené ou elle amenée à concevoir et assurer la mise en œuvre de chaînes de traitement structurant la collecte, la modélisation, l'analyse et la visualisation de corpus, en particulier textuels, ainsi que la publication en ligne de données scientifiques, en collaboration étroite avec des enseignants-chercheurs ou enseignantes-chercheuses, chercheurs ou chercheuses, personnels administratifs et techniques, doctorants ou doctorantes des établissements tutelles de la MSHB.

Activités

Activités principales

Mission 1 : Accompagnement de chercheurs dans leur projet de recherche / à ce titre l'agent a pour activités de :

- Assurer un accompagnement technique et méthodologique des projets de recherche soutenus par la MSHB, en proposant la structuration de chaînes de traitement adaptées aux corpus de données
- Accompagner la gestion et le traitement des données de recherche tout au long des projets
- Participer à la mise en œuvre de l'appel à projets humanités numériques

Mission 2 : Production, modélisation, analyse et visualisation de corpus, en particulier textuels / à ce titre l'agent a pour activités de :

- Réaliser des traitements et des analyses sur des corpus textuels
- Conseiller sur la production, la modélisation, l'analyse et la visualisation de corpus textuels
- Former les chercheurs et chercheuses à l'utilisation de logiciels d'analyse textuelle et les accompagner dans leur prise en main
- Collecter de jeux de données sur le web (web scraping) et conseiller sur les bonnes pratiques encadrant la collecte
- Réaliser une veille scientifique et méthodologique sur les méthodes et outils de traitement et d'analyse de données

Mission 3 : Dépôt de jeux de données (tous types) dans un entrepôt de données / à ce titre l'agent a pour activités de :

- Préparer un jeu de données en le structurant et en adaptant ses métadonnées à la norme utilisée par l'entrepôt
- Adapter un script pour le dépôt d'un jeu de données dans un entrepôt
- Nettoyer des jeux de données en vue de leur dépôt dans un entrepôt

Mission 4 : Relais en région de l'IR* Huma-Num / à ce titre l'agent a pour activités de :

- Informer et accompagner les communautés de recherche dans l'accès et l'utilisation des services d'Huma-Num
- Participer au réseau des correspondants Huma-Num et au réseau des modérateurs Nakala
- Former les chercheurs aux aspects techniques des services d'Huma-Num (API, scripts, etc.) et coanimer des ateliers avec les Datalabs de la MSHB

Mission 5 : Assurer la structuration du volet « traitement automatique du langage (TAL) » de la plateforme Humanités numériques/ à ce titre l'agent a pour activités de :

- Proposer des méthodes de traitement automatique des langues (TAL) pour l'analyse de corpus textuels, notamment l'analyse du discours, la reconnaissance d'entités nommées, la lemmatisation et l'analyse des sentiments.
- Assurer une veille sur les méthodes et les usages de l'intelligence artificielle (LLM, RAG, MCP) pour le traitement de données en SHS.

Activités secondaires

- Assurer des formations aux outils et aux méthodologies numériques employées, notamment en collaboration avec l'URFIST Bretagne-Pays de la Loire et dans le cadre des Datalabs de la MSHB
- Participer au comité d'organisation du séminaire Humanités numériques
- Participer au réseau des plateformes Scripto du Réseau national des MSH

- Documenter et déposer les méthodes de traitement développées dans le cadre des projets (via GitLab, par exemple)
- Encadrer des stagiaires participant aux projets scientifiques en traitement de corpus textuels

Activités annexes (en lien avec l'activité de la structure)

- Participer aux opérations de logistique de la structure lors de l'accueil d'événements scientifiques.

Niveau attendu

Titulaire d'un master 2 (de préférence en Humanités numériques, en Traitement automatique des langues ou en Linguistique/ Ingénierie numérique)

Connaissances requises

- Maîtrise du langage de programmation Python et de bibliothèques spécifiques à l'analyse et au traitement de données en sciences humaines et sociales, en particulier textuelles
- Maîtrise d'un ou de plusieurs logiciels d'analyse textuelle : Iramuteq, TXM, NVivo, QualCoder, Gargantext, etc.
- Maîtrise des méthodes d'interrogation et d'exploitation des API
- Connaissance des méthodes en TAL : recherche et extraction d'informations, annotation automatique, traitement statistique de corpus
- Connaissance du cycle de vie des données de recherche, des principes FAIR et des pratiques de stockage, de conservation et de dépôt dans des entrepôts de données (notamment Nakala, Cocoon et Ortolang)

Connaissances souhaitées

- Connaissance des protocoles de collecte de données en ligne (web scraping)
- Connaissance des standards de métadonnées (Dublin Core, OLAC)
- Connaissance des principes de modélisation et d'intégration de bases de données (relationnelles, graphes)
- Connaissance des méthodes et des enjeux de l'intelligence artificielle appliquée au traitement et à l'analyse de données en SHS (LLM, RAG, MCP, etc.)
- Connaissance et manipulation des formats de données structurées XML et JSON
- Connaissance des outils de gestion de versions du code informatique (Git)
- Connaissance des standards et technologies du web (HTML, CSS, JavaScript)

Compétences opérationnelles attendues

- Suivi de projet et accompagnement des équipes de recherche
- Rédaction et mise à jour d'une documentation fonctionnelle et technique
- Lecture de documentation en langue anglaise
- Animation de formations
- Capacités de travail en équipe et aisance relationnelle
- Capacités d'écoute et d'ouverture disciplinaire

Informations générales

Lieu de travail : MSHB – 2 avenue Gaston Berger, 35000 Rennes

Corps : ingénieur ou ingénieure d'études

Type de contrat : CDD de 11 mois avec une période d'essai de 1 mois

Date de prise de poste : début février 2026 (au plus tard)

Quotité de travail : temps complet

Branche d'activité professionnelle : BAP D - : Sciences Humaines et Sociales

Rémunération : selon la grille indiciaire des ingénieurs d'études (indice nouveau majoré 453)

Modalités pratiques du recrutement

La procédure de recrutement se déroulera en deux étapes : une sélection sur dossier et une audition pour les candidats ou les candidates dont les dossiers auront été retenus.

Le dossier de candidature devra être déposé sur le site de l'université Rennes 2, avant le mercredi 16 décembre (12h-heure de Paris) et devra impérativement être constitué d'un CV et d'une lettre de motivation.

Les auditions auront lieu à la Maison des Sciences humaines et sociales en Bretagne, à Rennes, début janvier.

Contact : Aline Benvegnu dos Santos, responsable de la plateforme Humanités numériques
aline.benvegnu-dos-santos [at] mshb.fr

09 novembre 2026