**Fiche de poste postdoc ArchivU**

**English version below**

**Contrat de recherche postdoctoral en TAL et/ou linguistique outillée (TAL)**

Le programme scientifique ArchivU annonce l’ouverture d’un contrat postdoctoral à temps plein sur 12 mois. Le postdoctorant ou la postdoctorante doit être titulaire d'un **doctorat en traitement automatique des langues ou en linguistique outillée** et avoir une expérience confirmée dans le **traitement de données textuelles numériques.**

**Contexte de la recherche**

Le projet ArchivU se donne pour but d’analyser les mutations de l’institution universitaire et du travail scientifique sur une diachronie courte, des années 1970 à nos jours, via l’analyse de deux genres textuels où se joue l’activité de l’université dans sa diversité : les comptes rendus des réunions du conseil d’administration et les rapports des laboratoires. Programme interdisciplinaire réunissant linguistes, historiens, philosophes, sociologues et conservateurs, ArchivU s’attache ainsi spécifiquement à des sources archivistiques encore largement inexploitées qui lui fournissent son corpus premier.

Le corpus ArchivU comprend l’intégralité des comptes rendus de l’université de Nanterre et de l’UPEC de 1971 à 2018 ainsi qu’une sélection de rapports de laboratoire de l’université de Nanterre depuis les années 1970. Ce corpus est destiné à être enrichi par trois niveaux de balisage : un **balisage physique** permettant de mettre en évidence la structure physique du texte ; un **balisage sémantique**, **en partie automatisé**, visant entre autres les entités nommées (noms de personnes, de lieux, dates, etc.) ; et une première ébauche de **balisage linguistique fin** des marqueurs caractéristiques des genres de discours étudiés (par exemple le discours rapporté et plus largement la polyphonie énonciative).

**Les missions proposées pour ce poste :**

Dans le cadre de la recherche ArchivU, le postdoctorant /la postdoctorante sera amené à proposer un projet de recherche en relation avec l’enrichissement et l’analyse du corpus ArchivU. La mission consistera à réaliser des tâches variées telles que :

-**Établissement, structuration, et nettoyage des données** (correction des sorties OCR et encodage XML) ;

- Prise en main/adaptation d’un **outil d’annotation manuelle multi-utilisateur** et de guides d’annotation

- Analyse des données avec une approche en **statistiques textuelles**

- Gestion et documentation des données et (méta)données sur la base d’un protocole préexistant à reproduire ;

- Mise en place d’**outils de détection automatique** à but d’enrichissement du corpus en rapport avec des marqueurs caractéristiques des genres de discours étudiés (**type fine-tuning de LLM** pour la détection et le balisage linguistique fin) ;

- Participation aux activités de publication et communication de l’équipe

- Formation des équipes aux méthodes et à l’utilisation des outils.

**Lieu de travail**

La personne recrutée travaillera à l’**université de Paris Est Créteil** et bénéficiera d’un bureau et d’un poste de travail au sein de l’UFR LLSH.

Début du contrat : entre le 1er octobre et le 15 décembre 2024

Rémunération : **2250,35 € net/mois**

**Profil recherché :**

- **doctorat en sciences du langage ou informatique**, spécialité traitement automatique des langues ou linguistique outillée  ;

- compétences en programmation **Python pour le TAL** (sklearn, Pytorch, Tensoflow…)

- expérience de l’annotation pour le TAL (gestion d’une campagne d’annotation)

-expérience dans le **traitement des fichiers XML**

- aptitudes à la rédaction et à la présentation de documents académiques ;

- **maîtrise du français** comme langue de travail principale ;

- capacité à travailler de manière indépendante, au sein d'une **équipe de recherche pluridisciplinaire**

- Capacité à prioriser les tâches, à prendre des initiatives

**Comment candidater :**

Les dossiers de candidature doivent comprendre :

- les coordonnées de contact (mail, téléphone) ;

- un CV actualisé avec une liste de publications ;

- une lettre de motivation intégrant des pistes de projet de recherches en lien avec la fiche de poste ;

- un texte académique et/ou projet de recherche jugé pertinent.

**Procédure de candidature** :

- La date limite de dépôt des candidatures est fixée au **20 juin 2024**.

- Le dossier de candidature complet doit être envoyé par e-mail à frederique.sitri@u-pec.fr avec l’intitulé « candidature Post-doc ARCHIVU »

- Les candidatures seront considérées comme complètes lorsque tous les éléments auront été reçus.

**Procédure de recrutement**

Une liste restreinte de candidats présélectionnés sur la base du dossier soumis sera invitée à un entretien qui aura lieu **entre le 1er et le 8 juillet 2024.**

**Contact**

**-** Les personnes intéressées sont invitées à prendre contact avec Frédérique Sitri ([frederique.sitri@u-pec.fr](mailto:frederique.sitri@u-pec.fr)) et Hugo Dumoulin (hdumoulin@parisnanterre.fr)

**Postdoctoral position in natural language processing or quantitative linguistics, M/F**

The ArchivU scientific program announces the opening of a full-time 12-month postdoctoral position. The postdoctoral researcher must hold a PhD in **natural language processing** or **quantitative linguistics**, and have proven experience in processing **digital textual data**.

**Scientific background**

The ArchivU project aims to analyze the transformations of the university as an institution over a short diachronic period, from the 1970s to the present day, through the evolution of two discourse genres: minutes of board meetings and laboratory reports. ArchivU is an interdisciplinary program bringing together linguists, historians, philosophers, sociologists and curators, and focuses specifically on the largely unexploited archival sources that provide its primary corpus.

The ArchivU corpus includes all the minutes of the University of Nanterre and UPEC from 1971 to 2018, as well as a selection of laboratory reports from the University of Nanterre since the 1970s.

It is intended to be enriched by three levels of tagging: physical tagging to highlight the physical structure of the text; semantic tagging, partly automated, such as, among other things, named entities (names of people, places, dates, etc.); and a first draft of fine-grained linguistic tagging of markers characteristic of the discourse genres studied (e.g. reported discourse and, more broadly, enunciative polyphony).

**Missions**

As part of the ArchivU research project, the postdoc will develop a research project in **quantitative linguistics and natural language processing** in relation with the main objectives of the project. The mission will involve various tasks such as:

- **Establishing, structuring and cleaning data** (correction of OCR output and XML encoding);

- Implementation/adaptation of a **multi-user manual annotation tool** and annotation guides.

- Data analysis using a **textual statistics** approach

- Data and (meta)data management and documentation based on a pre-existing protocol to be reproduced;

- Implementation of **automatic detection tools** for corpus enrichment in relation to markers characteristic of the discourse genres studied (**LLM fine-tuning**);

- Participation in the team's publication and communication activities

- Team training in methods and tools.

**Workplace**

**University of Paris Est Créteil**, UFR LLSJ ; office and computer are provided

Contract start date: between October 1 and December 15, 2024

Remuneration: **€2250.35 net/month**

**Profile required:**

- PhD in **language sciences or computer science**, specializing in **natural language processing or quantitative linguistics** ;

- **Python programming skills** for NLP (sklearn, Pytorch, Tensoflow...)

- experience of annotation for NLP (management of an annotation campaign)

-experience in **processing XML files**

- ability to write and present academic documents;

- **fluency in French** as main working language;

- ability to work independently as part of a **multidisciplinary research team**

- ability to prioritize tasks and take initiative

**How to apply :**

Applications must include :

- contact details (e-mail, telephone) ;

- an up-to-date CV with a list of publications;

- a cover letter including research project ideas in line with the job description;

- a relevant academic text and/or research project.

**Application procedure:**

- The deadline for applications is June 20, 2024.

- The complete application must be sent by e-mail to frederique.sitri@u-pec.fr with the heading "candidature Post-doc ARCHIVU".

- Applications will be considered complete once all elements have been received.

**Recruitment procedure**

A shortlist of candidates will be invited to an interview, which will take place between July 1 and 8, 2024.

**Contact**

- Frédérique Sitri (frederique.sitri@u-pec.fr) and Hugo Dumoulin (hdumoulin@parisnanterre.fr).