



Contrat post-doctoral en visualisation de données textuelles

Contexte:

Dans le cadre d'une collaboration avec des chercheurs de l'[EHESS](#), l'[ILC-CNR](#), le [Lattice](#), l'[Inalco](#), la [BNR](#) et le [BnF Data Lab](#), l'équipe-projet [ObTIC](#) - Sorbonne Université propose un contrat post-doctoral en visualisation de données textuelles, pour une durée de 12 mois renouvelable, à partir de septembre 2021 (Campus Jussieu).

Projet:

L'objectif du projet, baptisé *Textomining*, est de mettre à disposition de la communauté des outils numériques qui valorisent le patrimoine littéraire par l'analyse textuelle, la navigation et la visualisation. À partir d'un corpus numérisé et structuré de textes littéraires libres de droit (récits de voyage, romans, contes...), notre démarche consistera à repérer les noms de lieux et de personnages mentionnés dans les textes (les entités nommées), à analyser les modalités linguistiques autour de ces entités (émotions, sentiments, opinions...), et enfin à offrir des visualisations significatives de ces informations pour des lecteurs, chercheurs, enseignants ou étudiants.

La personne recrutée s'appuiera sur un travail déjà réalisé au sein de l'équipe autour du repérage des entités nommées et de leur analyse sémantique. À partir de textes français automatiquement pré-annotés (entités nommées, cadres spatio-temporels, structures dialogiques, déplacements spatiaux...), sa mission sera de proposer des modes de visualisation permettant de représenter des réseaux de personnages ainsi que leurs interactions et déplacements dans l'espace.

Ces représentations offriront des moyens de comparaisons chronologiques des œuvres et des auteurs à travers les lieux et les personnages, des moyens de tester des hypothèses, voire de faire émerger des connaissances nouvelles. Plus largement, ce travail permettra ultérieurement de créer des scénarios de navigation virtuelle autour d'un personnage ou d'un lieu donné, de relier les éléments identifiés à d'autres textes ou ressources en ligne (GeoNames, Wikidata, Gallica, IGN...). D'autres fonctionnalités éditoriales et de littérature augmentée pourront également être envisagées.

Profil recherché:

Une thèse de doctorat en informatique, en traitement automatique des langues ou en sciences humaines et sociales avec spécialisation en visualisation des données, en sémiologie graphique, en cartographie ou tout autre domaine similaire.

La personne recrutée doit avoir une bonne maîtrise des librairies de visualisation web (Javascript...) ou des applications SIG (système d'information géographique), une maîtrise d'un langage de script (de préférence Python).

D'autres connaissances de base restent nécessaires en traitement automatique des langues, dans les outils de repérage d'entités nommées (Stanza, Spacy...) et des technologies XML (TEI, XSLT...). Avoir des connaissances en statistiques est un plus.

La personne recrutée assurera le développement des solutions visuelles requises et leur intégration dans une chaîne de traitement existante. Elle travaillera en étroite collaboration avec les porteurs du projet et les différents partenaires. Pour certaines tâches de programmation, elle pourra être épaulée par des ingénieurs en informatique. Une journée d'étude conclura ce travail et une aide à la publication d'un article sera apportée.

Rémunération: salaire brut mensuel : environ 2500€ par mois.

Prise de fonction: septembre 2021.

Candidature:

Envoyer un CV et une lettre de motivation aux porteurs du projet avant le 19 juillet 2021:

Motasem Alrahabi: motasem.alrahabi@sorbonne-universite.fr

Glenn Roe: glenn.roe@sorbonne-universite.fr