

Offre de stage en Humanités Numériques au Labex OBVIL– Sorbonne Université

Contexte et principales missions :

Le stage se déroulera au sein de l'Observatoire de la Vie Littéraire (le labex [OBVIL](#)), dans le cadre du projet de la Très Grande Bibliothèque ([TGB](#)) qui vise à analyser et exploiter environ 130 milles documents XML ocrisés et non corrigés, issus des collections Gallica de la BnF et parus pour la grande majorité au XIX^e s. Après une phase de découverte de la base (structure, métadonnées, contenu...), la mission principale du stage consiste à expérimenter différentes applications TAL pour la reconnaissance automatique d'entités nommées (lieux et personnes). Le ou la stagiaire proposera des pistes pour améliorer la qualité des résultats à partir d'autres outils disponibles pour la modernisation et la correction semi-automatique de textes.

Compétences attendues:

Connaissance suffisante d'au moins un outil de reconnaissance d'entités nommées (Stanford NER, OpenNLP, Spacy...), connaissances en XML, Xpath et expressions régulières ; familiarisation avec les logiciels d'édition numérique de contenu (Oxygen...) ; maîtrise d'un langage de script (Python...).

Niveau de formation :

Master de formation TAL ou Humanités numériques.

Localisation :

Le stage se déroulera au Labex OBVIL, Maison de la Recherche, 28 rue Serpente, 75005 Paris.

Durée:

4 à 6 mois.

Début du contrat :

Dès que possible.

Gratification:

La gratification perçue est conforme à ce qui est prévu par l'université (soit environ 577.50€ par mois).

Candidature :

Pour toute question ou pour nous adresser votre candidature (curriculum vitae et lettre de motivation), merci d'écrire à motasem.alrahabi@sorbonne-universite.fr