

Proposition de contrat postdoctoral sur l'identification des types pédagogiques de documents issus de sites Web

Contexte : Établir des métadonnées sur des ressources, quelles qu'elles soient, est une tâche souvent considérée comme peu gratifiante, fastidieuse et consommatrice de temps. Elle est donc malheureusement, dans la majorité des cas ignorée lors de l'acquisition de la donnée et/ou de sa mise à disposition. De fait, ces données ne sont pas réellement valorisées car peu documentées et donc délaissées par les moteurs de recherche.

Par exemple, il existe un très grand nombre de ressources pédagogiques (ouvrages, vidéos, cours en ligne, ...) accessibles aux professeurs ou aux étudiants qui ne sont pas réellement utilisées voire totalement inconnues des utilisateurs potentiels. En effet, les auteurs (qui sont très souvent les professeurs eux-mêmes) n'ont pas été sensibilisés à cet aspect « métadonnées ». L'objectif de notre projet est de construire une base de métadonnées décrivant et pointant vers ces ressources, interrogeable à distance.

Objet du contrat : Le travail à réaliser dans le cadre du contrat post-doctoral portera principalement sur l'identification automatique du type pédagogique des ressources moissonnées sur des sites Web spécialisés. Il s'agit, à partir d'une liste de vocabulaire contrôlé et du contenu de ces ressources d'identifier le type pédagogique de chaque ressource afin de décrire le plus précisément possible les ressources concernées. Par exemple, à partir de la ressource "Jeu : Athènes et l'Acropole" provenant du site Web "jeux-historiques", le candidat devra identifier le type pédagogique de cette ressource et qui représente "jeu éducatif".

Pour cela, un traitement sémantique de ces contenus textuels est nécessaire en utilisant les différents outils de NLP ainsi que toutes autres mécanismes d'analyse de texte.

Ce post-doctorat sera réalisé au laboratoire ICube, en collaboration avec le LiLPa. Ces deux laboratoires rassemblent des chercheurs travaillant à la fois dans le domaine de l'informatique et dans divers domaines des sciences du langage, incluant la linguistique informatique. Il travaillera plus concrètement avec une post doctorante déjà à l'œuvre sur ce projet.

Candidature : Les candidat(e)s pour ce poste doivent être titulaire d'une thèse en informatique avec de très bonnes connaissances en traitement automatique des langues, fouille de textes et recherche d'information. La candidature doit comporter un CV détaillé (en français) incluant une liste de publications, une description de la thèse et son adéquation avec le sujet du post-doc ainsi qu'une lettre de motivation. Les coordonnées d'au moins deux référents sont à joindre.

Une très bonne compréhension de la langue française est indispensable : les ressources à modéliser et les publics concernés seront quasi exclusivement francophones.

Date souhaitée de début de contrat : janvier 2019 (au plus tard fin février 2019)

Durée : 6 mois

Salaire : de l'ordre de 2200 € net/mois négociable.

Contacts : Pierre Gancarski (gancarski@unistra.fr) et Delphine Bernhard (dbernhard@unistra.fr)

Téléphone : 03 68 85 45 76