

Contrat post-doctoral

Université Paris Est Créteil – Céditec (E.A. 3119) - Février 2012

Développement d'un outil dynamique d'exploration et d'observation des données textuelles multidimensionnelles (TextObserveur)

Profil recherché : Docteur-e en linguistique informatique, informatique ou TAL, intégration multimédia. Solides compétences en matière de développement informatique, avec une expérience dans la visualisation des données.

Rattachement : CEDITEC (E.A. 3119), Université Paris-Est Créteil Val de Marne, <http://ceditec.u-pec.fr/> ; chantier : Textopol <http://textopol.org/>

Lieu de travail: UPEC (Université Paris Est Créteil), Pyramide et CMC.

Type de contrat: contrat post-doctoral de 12 mois, à temps plein, à compter du 1^{er} septembre 2012.

Rémunération: 2000 euros nets mensuels.

Thème de recherche: Visualisation des données textuelles statistiques. (Chantier Textopol du Céditec)

Projet: *TextObserveur*, outil d'exploration et d'observation des données textuelles multidimensionnelles

Contexte : Une expérience importante et reconnue en matière d'outils logiciels pour la recherche et la formation en analyse des données textuelles à l'UPEC nous a conduits à développer le logiciel *TextObserveur*, dont plusieurs modules fonctionnent déjà. Fondé sur des préoccupations ergonomiques (navigation, interactivité, visualisations explicites, accessibilité pour l'utilisateur), cet outil propose des fonctionnalités originales sur le plan de la visualisation et du traitement dynamique des données et des résultats textométriques. Il rend possible l'intégration de données textuelles diversifiées dans un cadre multimédia.

Dans sa version finale, l'outil fournira aux chercheurs, aux apprentis-chercheurs et aux experts des sciences humaines et sociales (analyse de discours, linguistique de corpus, sciences de l'information et de la communication, lettres...) un dispositif complet d'aide à l'analyse mais aussi de formation à la recherche et à l'expérimentation. Il apportera un nouvel éclairage sur les corpus, répondra en temps réel aux questionnements expérimentaux portant sur les facteurs de la variation discursive. Il mettra à disposition des utilisateurs de nouvelles visualisations des textes, des corpus, des données issues du web, en s'adressant à l'ensemble des chercheurs en SHS intéressés par les données textuelles et multimodales, grâce à un système de lecture à plusieurs niveaux.

La première version de l'outil est disponible au téléchargement depuis mars 2011. Elle répond aux objectifs initiaux de la recherche : navigation simplifiée, dynamisation du mouvement, interactivité, représentation 3D. Une deuxième étape implémentera de nouvelles fonctionnalités textométriques dans le cadre multimodal et multilingue du web, et fera de *TextObserveur* l'interface du site de recherche et de la base textuelle Textopol, construits par nos soins depuis dix ans.

Missions du post-doctorant: Les travaux à la croisée du développement du web, de la visualisation, de la modélisation tridimensionnelle et de la statistique textuelle connaissent actuellement un développement considérable. Le post-doctorant devra dresser un état de l'art et une comparaison des expérimentations en cours, en matière de visualisation des données et de cartographies textuelles, telles que les travaux menés au sein du Médialab de Sciences Po Paris, ou les initiatives complémentaires du domaine telle que la plate-forme Cortext de l'IFRIS (Université de Paris-Est). Cette démarche comparative, évaluative et prospective permettra d'affiner les performances de *TextObserveur*.

Sur le plan des développements informatiques, il aura **pour mission de poursuivre le développement de l'interface** de navigation des données textuelles basée sur un modèle d'analyse factorielle des correspondances (3D et mouvement) tel que décrit aux adresses suivantes :

<http://textopol.org/contrib2010.php> <http://textopol.org/textobserver>

Il procédera en outre à l'installation des différents composants sur un serveur virtuel et au portage de la version locale de TextObservateur pour le web.

Profil détaillé: Titulaire d'un doctorat en linguistique informatique, informatique, TAL, ou sciences humaines, le candidat doit impérativement posséder de solides compétences en matière de développement informatique. Une spécialisation ou une expérience dans le domaine de la visualisation des données seront un atout pour le poste, tout comme la connaissance approfondie des travaux internationaux et des équipes travaillant dans le domaine. La qualité des publications fera l'objet d'un examen attentif.

Compétences techniques:

Données structurées en XML, base de données, développement web, développement en java (JOGL, 2D/3D en Java), C ou C++.

Connaissance d'un langage de script (Python, Javascript...), environnement de développement open source (SVN, Javadoc...Connaissance des concepts liés à ergonomie, maîtrise relative à la conception d'interfaces et interactions homme-machine.

Bonnes compétences ssh, ftp, environnement linux, installation de modules java sur un serveur...

Un intérêt pour le traitement automatisé des données textuelles est un atout pour le poste. La conférence JADT (<http://jadt.org>) est représentative de la communauté de recherche concernée.

Le candidat devra savoir s'intégrer dans un projet existant mais aussi faire preuve d'autonomie. Il devra en outre travailler en équipe et collaborer avec les différents acteurs du projet.

Modalités de candidature:

1. Dossier de candidature : CV détaillé, liste des publications et des activités scientifiques, lettre de motivation

Adresse d'envoi du dossier : UPEC – Céditec – A l'attention de Sonia Litaïem
La Pyramide
80 avenue du général de Gaulle
94000 Créteil cedex

Envoyer une **version électronique** du dossier à jean-marc.leblanc@u-pec.fr

Date limite d'envoi du dossier : 30 mars 2012

2. Entretien dans la première quinzaine d'avril

3. Soumission au Conseil Scientifique de l'UPEC pour validation (20 avril 2012)

Directeurs du Céditec : Caroline Ollivier-Yaniv yaniv@u-pec.fr, Dominique Ducard ducard@u-pec.fr

Contact et suivi du projet :

JM Leblanc jean-marc.leblanc@u-pec.fr
06.12.75.43.63
01.45.17.11.73

Responsable Textopol :

Pierre Fiala
fiala@u-pec.fr