



# Poste de Post-Doctorat



## Technologies anti-fraude big-data

Le laboratoire L3i, EA 2118, dans le cadre du projet SecurDoc, lance un appel à candidatures pour trois postes de post-doctorant en informatique dans le domaine de l'analyse d'images, de documents et la classification de documents.

*Durée : 12 mois*

*Date d'embauche : 1<sup>er</sup> octobre 2016*

*Salaire : 2100 € net / mois*

*Lieu de travail : dans les locaux du L3i à La Rochelle, France*

*Spécialité : Informatique / Traitement d'images / Classification / Analyse de documents*

### **Description du projet et du poste :**

Les travaux menés par le candidat se dérouleront au sein du L3i et s'inscriront dans le projet SECURDOC co-financé par l'entreprise Yooz d'une part, et des fonds Européens (FEDER) d'autre part.

Le laboratoire L3i, EA 2118 créé en 1993, représente la seule composante de recherche du domaine STIC à l'Université de la Rochelle associant les chercheurs de l'IUT de la Rochelle, et de la Faculté des Sciences en informatique. Dans le cadre de la politique quadriennale (désormais quinquennale) de l'université de la Rochelle, le L3i a été évalué A par l'AERES.

Le large déploiement des technologies numériques et la multiplicité des processus d'acquisition et de diffusion de l'information engendrent un développement rapide et diversifié des modes de production et de consommation de contenus numériques, ainsi qu'une croissance exponentielle de la volumétrie des données. Par ailleurs, l'avènement des dispositifs nomades interactifs augmente encore plus les problématiques de positionnement de l'utilisateur dans la gestion et la navigation au sein de contenus numériques.

Il s'agit, pour le L3i, de mettre en synergie les compétences établies dans le laboratoire afin d'aborder la problématique de la valorisation des contenus numériques sous un angle systémique. Cela revient, en particulier, à une exploitation croisée des compétences en matière d'applications interactives, d'indexation par le contenu, et de représentation de connaissances. Le laboratoire se structure autour de trois thématiques scientifiques (Ingénierie des connaissances, Analyse et gestion de contenus, Interactivité et dynamique des systèmes), toutes centrées sur la problématique de la gestion interactive et intelligente des contenus numériques.

Le projet SecurDoc, sélectionné dans le cadre des fonds FEDER de la région Nouvelle Aquitaine, rassemble l'entreprise Yooz et le laboratoire L3i. Ce projet vise à développer des prototypes logiciels complexes et innovants pour pouvoir étendre le périmètre d'un logiciel existant de capture de facture vers une gestion sécurisée de flux documentaires multicanal non-structurés. En particulier il s'agit de réaliser un prototype de laboratoire de détection de fraude de document numérique.

Plus précisément, l'objectif de ce travail sera de développer une fonctionnalité de fouille de contenus issus de sites publics du WEB afin de valider la cohérence des métadonnées extraites des documents. Partant des métadonnées standards contenues dans les documents, et de la masse de documents qui seront traitées, un contexte sémantique (dictionnaire de mots, d'adresses, ...) sera défini afin de confronter les données extraites de documents similaires ou fournies par un utilisateur, et celles issues de sources publiques fiables sur le web (listes publiques de société, ...).

Une expérience conduite par le L3i sur une problématique connexe montre le bénéfice de l'exploration du WEB pour vérifier des cohérences telles que l'adresse avec les personnes y résidants, le numéro de SIRET et le nom de la société, un logo et le nom de la société...

Par ailleurs, de nombreuses sources sont accessibles sur le WEB en France (exemple : l'état met à disposition des données fiscales), en Europe ou aux USA (exemple : les membres des organisations publiques).

Sur la base des savoir-faire existants nous souhaitons proposer un système de fouille de masse d'information générique. En particulier, dans ce projet, deux verrous technologiques seront à lever. Le premier concerne la découverte des sources à explorer pour valider des métadonnées car il n'est pas optimal de chercher en aveugle dans les masses d'informations disponible (Google par exemple). Sans recréer Google, un travail de pré-indexation orienté l'utilisation de nos métadonnées est nécessaire. Le second verrou réside dans l'interprétation des résultats du datamining avec des questions non résolus à ce jour: Comment la découverte de relations entre données est un indice de «vérité»? Quelle est la fiabilité des sources publiques? Comment gérer des réponses de fouille de données contradictoires?

#### **Qualifications spécifiques :**

Le candidat devra justifier d'une expérience de recherche dans au moins deux des domaines suivants :

- Traitement et analyse d'images
- Pattern recognition
- Traitement automatique de la langue
- Classification, machine learning

#### **Qualifications générales :**

- Maîtrise nécessaire d'un ou plusieurs langages de programmation (Java, Python, C/C++...)
- Très bonnes aptitudes au travail en équipe, une connaissance des méthodes Agile serait un plus
- Bonne aptitude à la rédaction d'articles scientifiques et maîtrise de l'anglais écrit et parlé
- Travail à l'interface avec le service R&D de l'entreprise partenaire

#### **Pour postuler :**

Les candidats à ce poste devront envoyer un CV et une lettre de motivation (les noms et coordonnées de références seraient un plus) à :

**quiquiqui [chez] univ-lr.fr ; mickael.coustaty [chez] univ-lr.fr ; vincent.poulaindAndecy [chez] yooz.fr ; saddok.kebairi [chez] yooz.fr**