

Offre de Stage en Informatique :

Développement de modules d'extraction d'information pour l'Enrichissement Sémantique d'Entités Culturelles

- **Poste** : Ingénieur d'étude / Développeur Informatique
- **Niveau minimum** : Bac+3 Informatique
- **Compétences** : Java, Maven, UML, RDF
- **Lieu du stage** : Campus de la Doua. Laboratoire LIRIS. UCBL Lyon 1.
- **Période du stage** : Juin/Juillet 2016 (selon organisation)
- **Gratification** : A discuter
- **Encadrants** : Joffrey Decourselle, Fabien Duchateau & Nicolas Lumineau
- **Date limite de candidature** : 20/05/2016

Contexte

Les technologies issues du Web Sémantique permettent aux institutions culturelles (bibliothèques, musées, archives, etc...) de s'équiper d'applications améliorant la gestion et réutilisation de leurs données informatiques (ex : les notices contenant les metadonnées des oeuvres culturelles). Toutefois, l'adoption de telles applications implique de pouvoir enrichir la sémantique des données initiales.

Ce stage s'inscrit dans un projet de recherche concret sur l'enrichissement sémantique d'entités culturelles ayant l'objectif de faciliter l'intégration des technologies du Web Sémantique dans les institutions culturelles. Les expérimentations faites dans le cadre du projet, se focalisent sur le traitement de notices bibliographiques issues de bibliothèques locales.

Cette expérience offre la possibilité de découvrir un contexte réel de recherche et développement impliquant des acteurs académiques et industriels et avec une forte dimension métier.

Objectifs

En partant de données RDF issues de catalogues de bibliothèques, l'objectif principal du stage consiste à intégrer des sources de données externes (e.g., DBPedia, Data.BNF) pour enrichir les entités en entrée. Ce processus est semi-automatique et implique des phases de validation et de fusion des entités. L'objectif secondaire est de proposer une interface web permettant de superviser le processus d'enrichissement et de manipuler les entités enrichies.

Travail à réaliser

- Réalisation de modules permettant l'extraction d'information depuis au moins deux sources de données comme DBPedia (<http://wiki.dbpedia.org/>) et Data.BNF (<http://data.bnf.fr/>)
- Création d'une interface de gestion/visualisation du processus d'enrichissement

Candidature

Emails : [joffrey.decourselle, fabien.duchateau, nicolas.lumineau]@liris.cnrs.fr

Adresse du laboratoire : LIRIS UMR 5205. Bâtiment Nautibus 43, bd du 11 novembre 1918, 69100, Villeurbanne

Envoyez un CV avec toutes les informations que vous jugez utiles aux trois emails.

Date limite de candidature : 20/05/2016