



Sujet de stage de Master

Projet DataLAC

« Intelligence artificielle et analyse d'un lac de données numériques de formats et contenus fortement hétérogènes pour l'archéologie »

Titre du stage : Apprentissage automatique pour la conception et l'analyse d'un lac de données

Lieu du stage : Laboratoire ERIC, Université Lyon 2 (campus Porte des Alpes, Bron, Rhône)

Encadrement : Jérôme Darmont et Sabine Loudcher

Durée : 5 à 6 mois à partir de février, mars ou avril 2023

Rémunération : Indemnités de stage légales (environ 550 € par mois / 35 heures par semaine)

Contexte du sujet de stage : DataLAC est un projet interdisciplinaire international financé par la Région AURA, qui vise à extraire des informations qualitatives (sémantiques) à l'aide de l'intelligence artificielle, afin de constituer un stock de métadonnées nécessaire à l'analyse de documents numériques stockés dans un lac de données. Le lac va contenir trois types de données :

- des bases de données structurées
- un corpus de transcriptions de carnets de fouille manuscrits concernant le site archéologique d'Ullastret en Catalogne
- des documents graphiques : croquis dans les carnets de fouille, dessins techniques (relevés de terrain, planches de profils de céramiques), photographies techniques (vestiges sur le terrain, couches stratigraphiques, planches d'objets), cartes et plans de diverses époques, voire des « scans » 3D.

L'objectif du projet DataLAC est l'extraction de métadonnées (concepts présents dans les documents ci-dessus) par des méthodes de transcription supervisée, non supervisée et de fouille de texte. Il s'agit de rechercher par similitudes aussi bien l'affectation directe de métadonnées descriptives (« ce document représente cette entité ») que la documentation de liens logiques entre documents (« ces deux documents représentent la même entité »). Toutes les métadonnées sont rassemblées et structurées dans un thésaurus bilingue catalan-castillan apparié avec des concepts en français.

Travail de stage : Un précédent stage a permis l'analyse de la structure des carnets de fouille, ainsi que la conception d'un thésaurus (vocabulaire structuré) et du schéma d'une base de métadonnées. Sur cette base, les tâches à mettre en œuvre par le stagiaire sont :

- l'instanciation de la base de métadonnées et la conception d'une interface de saisie des métadonnées
- la conception et la mise en œuvre du lac de données selon les principes FAIR¹, y compris la création de métadonnées de type liens entre les entités du lac, grâce à des méthodes d'apprentissage automatique (*machine learning*)
- la conception et la mise en œuvre d'une interface graphique d'analyse (exploration des données et *machine learning*) pour les archéologues.

Compétences souhaitées : bases de données, technologies *big data*, *machine learning*, programmation web.

Pour candidater : Envoyer CV, lettre de motivation et ainsi que notes de l'année universitaire précédente à adresser à jerome.darmont@univ-lyon2.fr et sabine.loudcher@univ-lyon2.fr avec en objet « Candidature stage informatique DataLAC » d'ici le **21 novembre 2022**. Les candidat·es retenu·es après examen des dossiers seront convoqué·es pour un entretien en visioconférence courant décembre.

1 Données Faciles à trouver, Accessibles, Interopérables, Réutilisables