

SUJET STAGE MASTER

Titre : Lac de données et intelligence artificielle pour la gestion des données multimédia : projet ESPHAISTOSS (projet du Chantier scientifique CNRS/Ministère de la Culture/Notre-Dame)

Mots-clés : Big data, lac de données, intelligence artificielle, métadonnées, histoire et données numériques

Nom des encadrants :

Jérôme Darmont (laboratoire ERIC), Sabine Loudcher (laboratoire ERIC), Mylène Pardoën (Maison des Sciences de l'Homme de Lyon Saint-Etienne)

Lieu :

Université Lyon 2, Campus Porte des Alpes, Bron / Réunion à la MSH-LSE (Lyon 7^e)

Durée :

4 à 6 mois à partir du 1er ou 15 septembre 2023

Rémunération :

4,05 euros par heure pour 35 heures de travail par semaine

Sujet :

ESPHAISTOSS est un projet pluridisciplinaire, qui vise l'étude et l'analyse des gestes et de la sensorialité de l'artisanat du bâti et l'artisanat d'art. Projet exploratoire, il se développe par modules qui s'imbriquent les uns aux autres avec des caractéristiques essentielles : les modules doivent être interopérables et ouverts, des outils au service de tous.

Les restitutions des gestes et de la sensorialité de l'artisanat (issues ou non de l'archéologie de terrain) prennent appui sur des captations visuelles et sonores synchronisées, une documentation hétérogène très importante. Or, ces outils sont utilisés par des chercheurs ayant des besoins spécifiques de gestion, mais également de traçabilité. La documentation nécessaire au projet ESPHAISTOSS représente de gros volumes de données hétérogènes et multimédia (aspects volume et variété des *big data*).

Dans le projet ESPHAISTOSS, le système de gestion de l'information de données multimédia devra à terme permettre le stockage, la modélisation, l'interrogation et l'analyse par des méthodes d'intelligence artificielle de toutes les données du projet : des données par nature multimédia et leurs métadonnées. Maîtriser l'hétérogénéité des données, gérer leur qualité et leur volume, garantir leur interopérabilité et leur accès efficace tout en les conservant dans leur forme originelle afin qu'elles demeurent des références utilisables peut s'envisager en exploitant le concept de lac de données (*data lake*).

Un premier stage recherche de master a permis d'explorer cette piste. Un état des lieux de l'existant du projet BRETEZ a permis de recenser toutes les données du projet avec des données dans une base de données, des documents textuels, des fichiers son, des fichiers image, une maquette dans Unity (un logiciel de développement de jeu vidéo), etc. La base de données a été entièrement restructurée et les premiers fondements du lac de données ont été posés.

L'objectif du deuxième stage recherche en 2023-2024 est de :

- poursuivre la construction du lac de données notamment avec le développement du système de gestion des métadonnées ;
- intégrer dans le lac les données provenant de la maquette vidéo ;
- de formaliser les besoins d'analyse des données du lac ;
- de concevoir les outils d'analyse à l'aide de méthodes d'intelligence artificielle tenant compte du contenu textuel, visuel et audio et de l'hétérogénéité des sources.

Compétences requises :

Le sujet de stage s'adresse à des étudiants en master ou en école d'ingénieurs d'informatique décisionnelle, de sciences des données. Des compétences en bases de données, en traitement des données massives, en technologies liées au Big Data ou en jeux vidéo seront particulièrement appréciées.

Contact :

Merci d'adresser, avant le 7 juillet 2023, votre candidature avec un CV, une lettre de motivation ainsi que vos notes de l'année universitaire en cours et de l'année dernière à jerome.darmont@univ-lyon2.fr, sabine.loudcher@univ-lyon2.fr et mylene.pardoen@cnrs.fr. Les candidats retenus seront convoqués pour un entretien au mi-juillet.