

SUJET STAGE MASTER

Titre :

Un lac de données pour la gestion des données multimédia du projet BRETEZ

Mots-clés :

Big data, lac de données, histoire et jeux vidéo

Noms de ou des encadrants :

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

Lieu :

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

Durée :

5 mois à partir de 1er mars ou du 1er avril 2020

Rémunération :

3,60 euros par heure pour 35 heures de travail par semaine

Sujet :

BRETEZ est un projet pluridisciplinaire qui vise la restitution visuelle et sonore de Paris dans la seconde moitié du XVIII^e siècle. 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 historiques urbaines (issues ou non de l'archéologie de terrain) prennent appui sur des moteurs de jeux vidéo ayant leurs propres caractéristiques liées à leur destination première, le jeu. Or, ces outils sont utilisés par des chercheurs ayant des besoins spécifiques de gestion, mais également de traçabilité. La documentation nécessaire au projet BRETEZ représente de gros volumes de données hétérogènes et multimédia (aspects volume et variété des *big data*).

Dans le projet BRETEZ, le module STRATEGIE (Système de TRAçabiliTE et de Gestion de l'information de données multimédia) devra à terme permettre le stockage, la modélisation, l'interrogation et l'analyse de toutes les données du projet, données par nature multimédia. 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 en 2018-2019 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 de master en 2019-2020 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 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 1^e ou 2^e année de master en informatique décisionnelle, en fouille de données ou en 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 15 janvier 2020, 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@wanadoo.fr. Les candidats retenus seront convoqués pour un entretien fin janvier.