

# Amélioration des modules d'une plateforme de textmining

HSM

Septembre 2022

## 1 Contexte général

Ce stage s'inscrit dans le cadre du projet ANR CROQUIS "Collecte, représentation, complétion, fusion et interrogation de données de réseaux d'eaux urbains hétérogènes et incertaines" financé par l'Agence Nationale de la Recherche (ANR-21-CE23-0004). Dans ce projet des chercheurs en Sciences de l'Eau et en Intelligence Artificielle unissent leurs efforts pour proposer de nouvelles méthodes pour la représentation, la complétion, la fusion, l'archivage, la réparation et l'interrogation des données hétérogènes décrivant les réseaux d'eau. Les informations textuelles utilisées dans le projet proviennent de la plateforme de textmining WEIR-P qui a été développée lors du projet "MeDo" (Megadonnées, données liées et fouille de données pour les réseaux d'assainissement-2018-2020), financé par la Région Occitanie-Pyrénées-Méditerranée à travers le dispositif "Recherche et Société(s) 2017"<sup>1</sup>. Le consortium scientifique du projet regroupait les UMRs HydroSciences Montpellier, Praxiling et TETIS ainsi que l'entreprise Berger-Levrault.

La plateforme WEIR-P combine des techniques de recherche d'information (RI) et d'extraction d'information (IE) adaptées à la langue française et au domaine des eaux usées. Le traitement en lui-même comporte 5 étapes : collecte de documents ; reconnaissance d'entités nommées ; extraction de relations sémantiques ; cartographie et visualisation de données ; exportation vers une base de données relationnelle et intégration dans un Système d'Informations Géographique.

## 2 Objectif

L'objectif du stage est de faire évoluer la plateforme WEIR-P.

Il s'agira plus particulièrement de :

- Améliorer le module de classification utilisé qui permet de trier les documents à analyser.
- Améliorer le liage des entités spatiales détectées par la plateforme avec les adresses incluses dans les bases de données BAN de l'IGN et NOMINATEM d'OSM.
- Interfacer la base de données PostgreSQL de MeDo avec QGIS via le plugin PostGis.
- Fusionner les données de la base PostGis MeDo avec celles provenant du Système d'Informations Géographique des gestionnaires du réseau.

## 3 Profil recherché

- Master 2 en informatique ou en Géomatique.

---

1. <http://webmedo.msem.univ-montp2.fr/>

- Outils et langages : une maîtrise de Python et de son écosystème est souhaitable.
- Capacité de travail dans une équipe pluri-disciplinaire.

## **4 Encadrement**

- Serge Conrad, UMR HydroSciences Montpellier, Université de Montpellier.
- Nanée Chahinian, UMR HydroSciences Montpellier, IRD.

## **5 Divers**

- Durée : 5 mois avec possibilité d'extension.
- Gratification : taux légal en vigueur
- Localisation : UMR HSM Montpellier
- Candidature : Envoyer un CV + relevés de notes des deux dernières années à :  
nanee.chahinian@ird.fr et serge.conrad@umontpellier.fr