

# Stage Master 2 ou école ingénieur

Conception et intégration d'une plateforme de visualisation de données textuelles hétérogènes

## Encadrement

**Jacques Fize** (UMR TETIS, CIRAD), **Mathieu Roche** (UMR TETIS, CIRAD), **Maguelonne Teisseire** (UMR TETIS, IRSTEA)

## Contexte général

Le stage se déroule dans le cadre du projet **SONGES** sur la **mise en correspondance de données textuelles massives et hétérogènes**. Dans ces travaux, nous élaborons des modèles de représentation de données ainsi que des mesures de similarité à partir d'indicateurs trouvés dans les textes (thématiques, spatiaux et temporels). L'objectif est d'organiser et valoriser des ensembles de données dans leurs dimensions hétérogènes et massives. Parmi les données exploitées, nous travaillons sur un ensemble de données produites dans le cadre du projet BVLAC, un projet mené par le CIRAD<sup>1</sup> qui promeut des techniques agricoles issues de l'agroécologie à Madagascar.

## Objectif

**L'objectif de ce stage est de développer une interface de visualisation** des liens (thématiques et spatiaux) entre documents d'un corpus. Cette interface devra permettre aux producteurs des données d'explorer et de valoriser ces corpus. Plus particulièrement, vous développerez une plateforme Web utilisant des bibliothèques dédiées telles que : *D3.js*, *Sigma.js*, *Topogram.io*, etc.

De façon plus précise, le stage sera décomposé en plusieurs étapes :

1. Appropriation du sujet (état de l'art, exploration des données)
2. Proposition de premières visualisations statiques des données à l'aide des bibliothèques disponibles sur Python ou R comme : *ggplot2*, *matplotlib*, *basemap*, *geopandas*
3. Conception de l'interface de visualisation
4. Stockage des données dans un SGBD<sup>2</sup>. La sélection du SGBD dépendra des besoins identifiés pour construire les différentes visualisations
5. Choix du **framework** (*Flask*, *Rshiny*,...) et des **bibliothèques Javascript** (*Sigma.js*, *Topogram.io*, *leaflet*, ...) nécessaires à l'implémentation de l'interface
6. Développement de l'interface
7. Analyse et évaluation des visualisations produites

## Compétences

- Langage de programmation : **Python** ou **R**
- Maîtrise de SGBD tels que *MariaDB*, *MongoDB*, *ElasticSearch* ou *PostgreSQL* (avec POSTGIS)
- Développement Web : **HTML/CSS** mais surtout **Javascript** (Connaissances en design d'IHM<sup>3</sup> souhaitées)

## Divers

- **Durée** : 6 mois
- **Gratification** : Taux légal en vigueur
- **Localisation** : Maison de la télédétection - Montpellier

## Comment candidater ?

Envoyer un CV ainsi que vos relevés de notes des deux dernières années à [jacques.fize@cirad.fr](mailto:jacques.fize@cirad.fr), [maguelonne.teisseire@irstea.fr](mailto:maguelonne.teisseire@irstea.fr), [mathieu.roche@cirad.fr](mailto:mathieu.roche@cirad.fr)

---

<sup>1</sup> Centre de coopération internationale en recherche agronomique pour le développement

<sup>2</sup> Système de Gestion de Base de Données

<sup>3</sup> Interface Homme-Machine