

**Type d'offre :** Offre en laboratoire

**Date de publication :** 24.01.24

**INRAE/MaIAGE**

# **STAGE M2 - Extraction d'Information sur les maladies transmises par vecteurs chez les plantes**

## **Informations générales**

**Type de contrat :** Stage

**Durée du contrat :** 6 mois

**Niveau d'étude :** Master/Ingénieur (BAC+5)

**Contact :**

[Claire Nedellec](#)

**Date de prise de poste :** ven 01/03/2024 - 12:00

**Métier :** Technicien

**Thématique :** IHM et visualisation données

## INRAE/MaIAGE :

L'Institut National de Recherche pour l'Agriculture, l'alimentation et l'environnement ([INRAE](#)) est un établissement public de recherche rassemblant une communauté de travail de 12 000 personnes, avec 272 unités de recherche, de service et expérimentales, implantées dans 18 centres sur toute la France. INRAE se positionne parmi les tout premiers leaders mondiaux en sciences agricoles et alimentaires, en sciences du végétal et de l'animal. Ses recherches visent à construire des solutions pour des agricultures multi-performantes, une alimentation de qualité et une gestion durable des ressources et des écosystèmes.

### Détail de l'offre (poste, mission, profil) :

#### Contexte de l'offre

Vous serez accueilli(e) au sein de l'[Unité MaIAGE](#) située à Jouy-en-Josas à ~15 km au sud-ouest de Paris. Outre Claire Nédellec, le stage sera co-encadré par Vincent Guigue (AgroParisTech), spécialistes des modèles de langue et Nicolas Sauvion (Phim, Montpellier) spécialiste des insectes vecteurs.

Vous travaillerez dans le cadre du projet *ANR BEYOND* sur la réduction des pesticides par l'anticipation des risques épidémiologiques des plantes. Vous aurez en charge le développement de nouvelles méthodes d'extraction automatique de relations biologiques impliquant des insectes vecteurs à partir de documents. L'enjeu sera d'extraire ces événements (relations n-aires) par des méthodes d'apprentissage profond (*deep learning, language model*) du traitement automatique de la langue (TAL) avec un nombre limité d'exemples produits manuellement et à l'aide de bases de connaissance externes.

Vous serez plus particulièrement en charge de :

- Réaliser un état de l'art des méthodes existantes d'extraction de relations n-aires et de citations.
- Proposer des extensions d'une de ces méthodes appliquées au sujet
- Intégrer la méthode dans le workflow ESV de TAL avec l'appui de l'ingénieur de l'équipe
- Evaluer les prédictions avec les mesures standards du domaine sur le corpus EPOP (*Epidemiomonitoring Of Plant*)

- Préparer un article en collaboration avec les co-encadrants en fonction des résultats obtenus.

## En savoir plus :

[Détails de l'offre \(INRAE\)](#) [Détails de l'offre \(Unité MaIAGE\)](#)

## Profil et compétences recherchés

- **Formation recommandée** : Master 2 en Informatique, ou 3è ingénieur, spécialité Traitement Automatique des Langues et/ou Apprentissage automatique ;
- **Connaissances souhaitées** : programmation en Python et/ou Java., anglais ou français courant ;
- **Expérience appréciée** : expérience de deep learning, pratique du TAL et/ou utilisation de la bibliothèque HuggingFace ;
- **Aptitudes recherchées** : Intérêt pour les applications de l'informatique à la biologie et le travail interdisciplinaire.

**Date limite pour postuler** : mer 31/01/2024 - 12:00

**URL de l'offre** : <https://www.dataia.eu/sites/default/files/Stage%20IEV.pdf>

**Lien vers l'offre sur le site dataia.eu** :<https://da-cor-dev.peppercube.org/node/845>