

PROPOSITION DE STAGE MASTER 2 IEV 2024

Extraction d'Information sur les maladies transmises par vecteurs chez les plantes

Information extraction on vector-borne diseases in plants from the literature

Lieu : unité MaIAGE, Centre de recherche INRAE de Jouy en Josas,

Durée : 5 à 6 mois entre février et août 2024

Contacts :

- Claire Nédellec (équipe Bibliome, unité MaIAGE, [INRAE](#)), Jouy-en-Josas.
claire.nedellec@inrae.fr
- Vincent Guigue (équipe EkINOCs, unité MIA-Paris-Saclay, AgroParisTech) Palaiseau.
vincent.guigue@agroparistech.fr
- Nicolas Sauvion (équipe FORISK, unité PHIM, INRAE), Montpellier.
nicolas.sauvion@inrae.fr

Sujet du stage

Le projet porte sur l'extraction automatique de relations biologiques à partir de documents. Il ciblera en priorité trois espèces particulières de psylles (insectes piqueurs suceurs) vecteurs de bactéries responsables de maladies des arbres fruitiers. Les événements représentant les interactions biologiques entre microbe, insecte, plante et leurs lieux et dates d'observation sont dénotés dans les textes scientifiques par des formulations complexes variables qui portent fréquemment sur plusieurs phrases. L'enjeu sera d'extraire ces événements (relations n-aires) par des méthodes d'apprentissage profond (*deep learning, language model*) avec un nombre limité d'exemples produits manuellement et à l'aide bases de connaissance externes.

Profil

- Master 2 en INFORMATIQUE orienté Traitement Automatique des Langues et/ou Apprentissage automatique
- Expérience de *deep learning*
- Expérience en TAL et/ou utilisation de la bibliothèque HuggingFace
- Maîtrise de l'anglais ou français courant.
- Compétences techniques requises : Python et/ou Java
- Intérêt pour les applications en biologie et le travail interdisciplinaire.

En savoir plus : <https://maiage.inrae.fr/sites/default/files/2023-12/Stage%20IEV.pdf>