

## **PsylVe : un cadre conceptuel et pratique de Text Mining pour l'extraction d'information de données d'occurrence de vecteurs de pathogènes de plantes.**

Les maladies dues aux phytopathogènes transmis par les insectes ont un effet négatif important sur l'agriculture mondiale. Un moyen efficace d'anticiper les épidémies est d'inférer des cartes de risque d'introduction et de propagation des vecteurs à partir de données d'occurrence connues. Mais la compilation manuelle de ce type de données est longue et laborieuse, en particulier en raison de l'augmentation croissante des données accessibles. Pour résoudre ce problème, ce stage visait à faciliter la tâche des chercheurs et des veilleurs des plateformes d'épidémiosurveillance en utilisant des approches permettant d'automatiser l'extraction d'informations relatives aux vecteurs à partir de la littérature. Dans ce but, nous avons développé PsylVe, une solution initialement ciblée sur les vecteurs de psylles qui englobe des techniques de collecte de documents, de traitement du langage naturel (NLP) et de représentation des connaissances (KR). PsylVe comprend un pipeline NLP fonctionnel et une méthodologie entièrement documentée. Nous avons effectué une évaluation quantitative (précision, rappel et score F1) et qualitative (six critères qualitatifs pour l'évaluation des pipelines de fouille de textes) des résultats obtenus avec PsylVe et nous les avons comparés à un ensemble de données compilées manuellement d'observations sur le psylle *Cacopsylla pruni*, responsable de la propagation d'un phytoplasme dans les vergers d'arbres fruitiers en Europe. Dès sa genèse, nous avons conçu PsylVe pour qu'il soit transférable à d'autres vecteurs de maladies végétales, voire des maladies humaines et animales. Nous avons également conçu une application pour l'extraction de textes à partir de documents PDF et une ontologie formelle originale qui permet la représentation des données et des connaissances sur les maladies à transmission vectorielle.