

INTERNSHIP PROPOSAL MASTER 2 BIOCONTROL 2026

Towards Effective Biocontrol: Automatic Extraction of Experimental Results from Scientific Literature

Duration : 5 to 6 months between February and August 2026

Location: MaIAGE, INRAE research center in Jouy en Josas,

Gratification: Approximately 600 euros per month

Keywords : Natural Language Processing, machine learning, language models, information extraction, plant protection, biosolution

Internship supervisors

- Claire Nédellec (Bibliome team, MaIAGE lab, [INRAE](https://www.inrae.fr)), Jouy-en-Josas.
claire.nedellec@inrae.fr
- Thomas Presseccq, société [Syndev](https://www.syndev.fr). thomas.presseccq@inrae.fr

Profile

- Master's degree in Computer science with a focus on Natural Language Processing and/or Machine Learning
- Experience in language model training and application
- Experience in NLP and/or use of the HuggingFace library
- Fluency in English or French.
- Required technical skills: Python and/or Java
- Interest in biology and interdisciplinary work.

Application

CV, motivation letter, L3, M1 and M2 grades should be send to claire.nedellec@inrae.fr

Examined on a real-time basis

Subject

Context

This internship is part of the OAD Success project - Decision Support Tool to Promote the Use and Success of Biosolutions for Plant Health. Improving the use of biosolutions as an alternative to synthetic pesticides requires a good understanding of the factors that influence their effectiveness. Biosolutions contain active molecules, beneficial microbial flora, and plant extracts that are tested on different plants or parts of plants (roots, leaves). Application methods, such as the dosage of active molecules, frequency, or formulation (granules, spray), have a decisive effect. Biosolutions are tested in highly controlled environments or in conditions close to those found in the field, with contrasting results. The effects are measured in terms of reduction in pests or reduction in damage to plants. The scientific literature contains a large number of descriptions of experiments with biosolutions and their results under different conditions.

Example:

*In the present study, bacterial endophytes **B. subtilis GLB191** and **B. pumilus GLB197** originally isolated from **grapevine leaves** demonstrated their biocontrol potentials against **downy mildew disease in leaf disks and field conditions**. [Zhang et al ., *Crop protection*, 2017].*

The aim of the project is to automate the collection and formalization of this information, which has until now been done manually. It will be produced in a structured and standardized format so that it can be entered and compared with the information obtained manually. The results of the internship will be used by the decision-making support application developed by SYNDEV with the aim of helping farmers reduce their use of synthetic inputs.

Objective

The research internship focuses on the development of automatic methods for extracting information from documents in English on biosolutions, their methods of application, experimental conditions, and their effectiveness. This information on the various experimental parameters and their values is expressed in different forms in the documents, texts, tables, and figures. The project already has high-quality information extracted manually from the documents.

The aim of the internship will be to develop methods for extracting information from scientific texts based on deep learning with few or no examples (zero-few-shot deep learning). In collaboration with the project's biologists, the most relevant types of entities will be selected and different language models will be tested and adapted according to the nature of the entities (proper noun, variable compound word, measurement) and their context (text, table). The information previously acquired manually will be used to train the models and evaluate them. The models considered by the internship will take into account realistic constraints of frugality, and larger models may be tested for comparison purposes.

The conclusions of the internship will include an analysis of the remaining obstacles and promising solutions to be explored further.