

4 Stages de Master 2/Ingénieur - 2027

Extraction de graphes de connaissances au service de l'agro-science

Présentation INRAE

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. Premier organisme mondial spécialisé sur l'ensemble agriculture – alimentation – environnement, INRAE joue un rôle clé pour accompagner les transitions nécessaires face aux grands défis planétaires.

Environnement de travail

Projet DIGIT-BIO-SEED

Le projet vise à utiliser les ressources sémantiques du domaine végétal et de l'alimentation, en valorisant son apport à travers quatre applications centrées sur les plantes et la nutrition. Son objectif principal est d'améliorer l'annotation sémantique des connaissances sur les plantes et les aliments, afin de développer un outil pour faciliter l'inférence d'associations entre les composés chimiques et les concepts biologiques associés.

Nous proposons 4 stages dans des unités de recherche localisées dans différentes régions de France (Rennes, Bordeaux et Toulouse), composant l'infrastructure nationale (MetaboHUB).

UMR 1332 BFP (Bordeaux)

L'unité de recherche Biologie du Fruit et Pathologie joue un rôle majeur dans la recherche en biologie végétale en région Nouvelle-Aquitaine. L'équipe MéTA est une équipe multidisciplinaire qui s'intéresse au fonctionnement des plantes, notamment celui du fruit. Nous recherchons des marqueurs métaboliques associés au métabolisme central mais également au métabolisme redox en raison de leur implication dans l'adaptation aux contraintes environnementales.

UMR 1349 IGEPP (Rennes)

L'Institut de Génétique, Environnement et Protection des Plantes est un laboratoire de recherche en protection des plantes au sein duquel la plateforme P2M2 fournit un service analytique en chimie métabolomique. Cette unité mobilise ses projets de recherche pour explorer et comprendre chez les Brassica cultivées, les attributs biologiques de la vigueur végétative, de l'allocation des ressources et de la qualité des graines dans le contexte de l'acclimatation aux changements climatiques.

UMR 1268 BIA (Nantes)

L'unité de Biopolymères Interactions Assemblages est un laboratoire de recherche dédié à l'étude des constituants des matrices agro-alimentaires. Au sein de cette unité, l'équipe PRP, intégrée à la plateforme P2M2, contribue plus spécifiquement à la caractérisation fine des profils métaboliques et des biopolymères par des approches analytiques avancées, en appui aux projets visant à concevoir des aliments plus durables et de meilleure qualité.

UMR 5546 LRSV (Toulouse)

Le Laboratoire de Recherche en Sciences Végétales abrite la plateforme d'analyse métabolomique Métatoul-AgromiX, qui développe des bases de données comptant plusieurs millions de spectres ainsi que des outils originaux de traitement de données LC-MS pour l'analyse de matrices végétales complexes, incluant notamment des espèces d'intérêt agronomique et médicinal.

Problématique

La littérature scientifique recèle une masse d'informations sur le rôle déterminant de nombreux métabolites dans les produits végétaux, influençant à la fois la résistance aux maladies, les interactions avec l'environnement et les propriétés organoleptiques. Ces connaissances sont cruciales pour comprendre les mécanismes qui façonnent les caractéristiques des plantes, leur réponse aux stress et leur potentiel de transformation.

Le domaine végétal dispose par ailleurs d'ontologies de référence (Plant Ontology, Trait Ontology, Plant Experimental Conditions Ontology issues de Planteome) et, pour la transformation des bioressources, de l'ontologie TransformON développée à l'INRAE. Dépasant leurs usages actuels, ces ressources peuvent servir à annoter la littérature scientifique de façon standardisée, ce qui facilite l'intégration et la réutilisation des connaissances.

L'enjeu central est désormais d'exploiter ces annotations, intégrées dans des graphes de connaissances, pour rendre la veille scientifique plus efficace et en automatiser une partie dans des contextes applicatifs.

Missions - Objectifs du stage

FORVM DiseaseChem (<https://forum-webapp.semantic-metabolomics.fr>) est un graphe de connaissances biomédical, robuste et publié, qui relie des composés chimiques, en particulier des métabolites, à des maladies et à d'autres concepts ontologiques. Le graphe *FORVM Plants* reprend cette méthodologie et l'applique au domaine des sciences du végétal et de l'alimentation.

Le stagiaire devra exploiter des sous-graphes dérivés du graphe *FORVM Plants* spécifiques qui relient des composés chimiques à des concepts ontologiques pertinents pour les sciences végétales et/ou l'alimentation. Il travaillera sur un des quatre cas d'étude sélectionnés : 1) *Développement du fruit de tomate/Bordeaux* ; 2) *Stress biotique et abiotique/Rennes* ; 3) *Interactions protéines-polyphénols/Nantes* ; 4) *Effet de l'environnement/thérapeutique sur la composition du cannabis/Toulouse*. Il devra partager les avancées de son travail de façon hebdomadaire en visioconférence avec les partenaires du projet *DIGIT-BIO-SEED*.

Résultats attendus

- Extraction et formalisation d'un ou plusieurs sous-graphes de connaissances spécifique à l'un des cas d'étude à partir de FORVM Plants ;
- Mise en place d'un prototype d'analyse expérimentale (par exemple un jeu de données métabolomique) contextualisée par ce sous-graphe ;
- Validation des sorties en montrant comment les connaissances extraites peuvent aider à interpréter les résultats.

Compétences à développer en web sémantique

- Comprendre les principes du Web de données : *RDF*, *URIs*, triples, graphes de connaissances.
- Connaître les principaux standards du *W3C* : *RDF*, *RDFS*, *OWL*, *SPARQL*.
- Maîtriser les notions de graphes de connaissances et de données liées (Linked Data).
- Savoir manipuler *RDF* dans différents formats : *Turtle*, *RDF/XML*, *JSON-LD*, *N-Triples*.
- Savoir installer et utiliser un triplestore (par exemple *Virtuoso*, *Apache Jena/Fuseki*).
- Manipuler des graphes *RDF* avec des bibliothèques adaptées (par exemple *RDFLib* en *Python*).
- Connaître et exploiter les ontologies du domaine des plantes, notamment *PLANTEOME* (<https://planteome.org/>).

Formations et compétences recherchées

Stage de fin d'étude Master 2 / Ingénieur (janvier à juin 2027)

Nous recherchons un(e) stagiaire **niveau Master 2 / Ingénieur**, motivé(e) par les environnements scientifiques et techniques complexes, avec une forte appétence pour la multidisciplinarité. Le/la candidat(e) évoluera à l'interface de plusieurs domaines (informatique, ingénierie de la connaissance, biologie des plantes ou agroalimentaire, selon le contexte), et devra faire preuve de curiosité intellectuelle, de capacité d'adaptation et d'un goût prononcé pour le travail collaboratif entre les 4 stagiaires recrutés. Une sensibilité pour la biochimie ou pour les bases de données de composés chimiques sera particulièrement appréciée.

Contacts :

- Bordeaux : UMR BFP - Cécile Cabasson - cecile.cabasson@u-bordeaux.fr
- Rennes : UMR IGEPP - Younes Dello - younes.dello@inrae.fr
- Nantes : UMR BIA - Magalie Weber - magalie.weber@inrae.fr
- Toulouse : UMR LRSV - Guillaume Marti - guillaume.marti@utoulouse.fr
- Chef de projet DIGIT-BIO-SEED - Olivier Filangi - olivier.filangi@inrae.fr