

Offre de Stage

Construction de graphe de connaissances à partir de bases de données relationnelles

Lieu: INRAE Centre Occitanie Montpellier, UMR MISTEA, bâtiment 21, campus supagro, 2 Place Pierre Viala, 34060 Montpellier

Profil étudiant : Ingénieur ou M2 en informatique, composante génie logiciel ou système d'information, avec un fort intérêt pour l'agriculture ou l'environnement et la modélisation ou Ingénieur agronome ou M2 avec une spécialisation en science de données ou agroTIC.

Durée : 6 mois à temps plein

Equipe INRAE UMR MISTEA: Catherine Roussey, Farzaneh Kazemipour, Pascal Neveu

Equipe INRAE UMR LEPSE: Pierre Martre, Llorenç Cabrera-Bosquet

Discipline/mots clés : Modélisation (modèles de culture), Expérimentation agronomique, Phénotypage, Ressources sémantiques, Schéma de données, Génie logiciel, ontologie.

Sujet :

Les défis de la sécurité alimentaire, de la réduction des impacts environnementaux et de l'adaptation de l'agriculture au changement climatique nécessitent de repenser l'agriculture. Relever ces défis et éclairer les décisions nécessite de nouvelles expériences. Ces expériences produisent beaucoup de données variées (mesures de capteurs, observations humaines, enquêtes, satellites, drones, etc). Pour que ces données soient exploitables d'autres types de données sont nécessaires comme la description des ressources génétiques, événements météorologiques (gels, inondations, sécheresses). Les systèmes d'information doivent gérer cette diversité et se connecter à d'autres systèmes de gestion de données, tout en répondant aux exigences d'une science ouverte et reproductible. Les technologies du Web Sémantique (RDF, OWL, SKOS) sont une solution prometteuse au partage de données. **Le sujet de stage proposé vise à soutenir l'intégration de données continues issues d'expérimentations agronomiques.** L'objectif du stage est de structurer les données dans des **graphes de connaissances** fondés sur un réseau d'ontologies du domaine de l'expérimentation agronomique (OESO et AgrO). Une ontologie est un schéma de données orienté graphe partagé sur le Web à l'aide des technologies Web Sémantique (W3C). Un graphe de connaissances contient les faits instanciant les composants d'une ontologie et permet d'introduire de l'intelligence dans les applications.

La communauté de recherche en expérimentation agronomique a développé un format standard d'échange de données, intitulé ICASA [White et al, 2013] qui a doit évoluer pour répondre aux grands défis de l'agriculture. Les équipes de recherche de MISTEA et du LEPSE développent un nouveau schéma conceptuel orienté base de données relationnelles pour clarifier le format ICASA. Ce nouveau schéma est construit en utilisant le langage UML avec l'outil draw.io. **L'objectif du stage est de traduire ce schéma conceptuel orienté base de données relationnelles vers un graphe de connaissances.**

Le stagiaire sera en charge de :

- l'étude du schéma relationnel produit et de sa procédure de mise à jour (Qui, Quand, Comment, avec quoi)
- l'analyse des règles de transformations des schémas de bases de données relationnelles vers une ontologie [Bouazzouni et Jonquet, 2021] [Ben Mahria et al, 2021] [Mosca et al, 2023].
 - Déterminer les règles qui peuvent s'appliquer automatiquement de celles qui nécessitent une intervention humaine.
- la prise en main de l'outil draw.io et des fichiers sous jacent
- la mise en oeuvre du langage CHOWLK et de l'outil associé CHOWLK convertir basé sur draw.io (<https://github.com/oeg-upm/Chowlk>) [Chávez-Feria et al, 2022]
- la conception d'une procédure semi automatique de mise à jour entre le diagramme UML relationnel et diagramme CHOWLK de l'ontologie.
- l'implémentation d'un outil de traduction UML relationnel vers CHOWLK dans draw.io.

Pour toutes ces activités, le stagiaire devra interagir avec l'équipe d'agronomes. En fonction de l'appétence de la personne recrutée, le sujet peut prendre un aspect plus méthodologique ou plus implémentation informatique. La personne recrutée bénéficiera d'un environnement scientifique interdisciplinaire riche et dynamique porté par plusieurs projets internationaux. Ce stage est soutenu par #Digitag (l'institut d'agriculture numérique) et pourra être suivi d'un contrat.

Compétences requises : Savoir faire en modélisation informatique langage UML (logiciel draw.io), en modélisation de schéma conceptuel (bases de données relationnelles) et/ou modélisation graphes de connaissances. Le candidat devra avoir un goût prononcé pour l'innovation.

Indemnités et Conditions : (financement Digitag) environ 550 euros par mois. Ordinateur portable de fonction et logiciels mis à disposition par MISTEA.

Contacts et candidatures: Votre candidature sera à envoyer à catherine.roussey@inrae.fr et farzaneh.kazemi-pour@inrae.fr. La candidature doit comporter un CV, une lettre de motivation et les coordonnées (email et téléphone) d'une ou deux personnes référentes. Suite à la réception de votre CV nous vous proposerons un entretien en visio conférence.

Références

- [White et al, 2013] J. W. White, L.A. Hunt, K. J. Boote, J. W. Jones, J. Koo, S. Kim, C. H. Porter, P. W. Wilkens, G. Hoogenboom. Integrated description of agricultural field experiments and production: The ICASA Version 2.0 data standards. *Computers and electronics in agriculture*, 2013, vol. 96, p. 1-12. <https://www.ars.usda.gov/ARUserFiles/20200500/White%20et%20al.%20CEA2013.pdf>
- [Chávez-Feria et al, 2022] S. Chávez-Feria, R. García-Castro, and M. Poveda-Villalón. Chowlk: from UML-Based Ontology Conceptualizations to OWL. In *The Semantic Web: 19th International Conference, ESWC 2022, Hersonissos, Crete, Greece, May 29 – June 2, 2022, Proceedings*. Springer-Verlag, Berlin, Heidelberg, 338–352. https://doi.org/10.1007/978-3-031-06981-9_20
- [Bouazzouni et Jonquet, 2021] S. Bouazzouni, C. Jonquet. L'ontologie E-Phy, une base de connaissances pour le catalogue des produits phytopharmaceutiques autorisés en agriculture en France. IC 2021 - 32es Journées francophones d'Ingénierie des Connaissances, 2021, Bordeaux, France. pp.105-112. (emse-03260693)
- [Ben Mahria et al, 2021] B. Ben Mahria, I. Chaker, A. Zahi. A novel approach for learning ontology from relational database: from the construction to the evaluation. *Journal of Big Data*, 2021, vol. 8, no 1, p. 25.
- [Mosca et al, 2023] R. Mosca, M. De Santo, R. Gaeta. Ontology learning from relational database: a review. *Journal of Ambient Intelligence and Humanized Computing*, 2023, vol. 14, no 12, p. 16841-16851.