

Titre

Aide à la correction d'anomalies dans des données multidimensionnelles et multirelationnelles sur l'agroécologie en santé animale et végétale

Résumé

Pour réussir la transition agroécologique, les producteurs ont besoin de disposer de connaissances sur des alternatives aux techniques agricoles classiques. Cependant, en préalable à l'utilisation d'une base de connaissances (BC) par des producteurs et des experts scientifiques, celle-ci doit être corrigée de ses anomalies. Le contexte de ce stage est la BC Knomana [Silvie et al., 2021], qui rassemble 48000 descriptions d'utilisation de plantes à effet pesticide et antibiotique, et vise à proposer des préparations à base de plantes en remplacement des produits chimiques de synthèse. Des dictionnaires permettent déjà de corriger les valeurs pour ses 31 types de données. Par contre, la vérification de la correction et de la cohérence des données est trop complexe pour être réalisée manuellement. Par exemple, une incohérence entre la plante pesticide, le système protégé (e.g. culture de maïs), le bioagresseur (e.g. insecte) et la localisation géographique suffit pour induire en erreur un producteur. La technique appelée Exploration d'Attributs (EA), développée par l'Analyse de Concepts Formels, permet de détecter et de corriger ces anomalies [Saab et al., 2022] en exprimant chaque connaissance sous forme d'une règle d'implication. Les règles sont présentées aux experts qui les valident ou les invalident afin de mettre la BC dans un état cohérent.

L'objectif du stage est de développer un prototype logiciel de détection et de correction des anomalies pour des données multidimensionnelles et multirelationnelles. Ce prototype permettra de manipuler les données et les types de données, puis d'interagir avec la librairie FCA4J, pour le calcul des règles, et le logiciel RCAvizIR, développé avec le soutien de #Digitag (stages de Master en 2022 et 2023) pour les présenter dans un ordre facilitant le travail de correction par les experts.

Lieu

Laboratoire d'Informatique, Robotique et Microélectronique de Montpellier (LIRMM)

Encadrement

Alexandre Bazin (alexandre.bazin@lirmm.fr), LIRMM (WEB3)

Marianne Huchard, LIRMM (MAREL)

Pierre Martin, CIRAD (AIDA)

Pascal Poncelet, LIRMM (ADVANSE)

Arnaud Sallaberry, LIRMM (ADVANSE)

Compétences recherchées

Intelligence artificielle, Fouille de données, Javascript

Mots clés

Agroécologie, Usage des plantes, Analyse de Concepts Formels, Règle d'implication multidimensionnelle, Visualisation, Détection et Correction d'anomalies

Discipline principale du projet

Informatique, Extraction de connaissances, Visualisation

Discipline secondaire du projet

Sciences de la Vie et de l'Environnement

Bibliographie

* Pierre Accorsi, Mickaël Fabrègue, Arnaud Sallaberry, Flavie Cernesson, Nathalie Lalande, Agnès Braud, Sandra Bringay, Florence Le Ber, Pascal Poncelet, Maguelonne Teisseire. HydroQual: Visual Analysis of River Water Quality. Proceedings of the IEEE Symposium on Visual Analytics Science and Technology (VAST), pp. 123-132, 2014.

* Daniel A. Keim, Gennady L. Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, Guy Melançon. Visual Analytics : Definition, Process, and Challenges. Information Visualization - Human-Centered Issues and Perspectives. LNCS 4950, Springer 2008, p. 154-175. * Tamara Munzner. Visualization Analysis & Design. CRC Press, A K Peters Books, 2014. * Roberto Tamassia, Handbook on Graph Drawing and Visualization. Chapman et Hall / CRC, 2013.

* Michael Sedlmair, Miriah D. Meyer et Tamara Munzner. Design Study Methodology: Reflections from the Trenches and the Stacks. IEEE TVCG 18(12): 2431-2440, 2012.