

Stage de Master 2 Recherche ou Professionnel
Master Informatique ou Bioinformatique
(English version below)

Navigation dans les règles d'implication multidimensionnelles
sur l'agroécologie pour l'aide à la décision en santé animale et végétale

Le stage est financé par l'Institut de convergence [#Digitag](#)

Durée **6 mois**

Début : **Février 2023**

Laboratoire [LIRMM](#)

[Gratification de stage](#) (cas des organismes publics)

Encadrants du stage

Marianne Huchard ([LIRMM <marianne.huchard@lirmm.fr>](#)), Alexandre Bazin ([LIRMM <alexandre.bazin@lirmm.fr>](#)), Pierre Martin ([CIRAD <pierre.martin@cirad.fr>](#)), Pascal Poncelet ([LIRMM <Pascal.Poncelet@lirmm.fr>](#)), Arnaud Sallaberry ([LIRMM <Arnaud.Sallaberry@lirmm.fr>](#))

Mots clefs

Agroécologie, usage des plantes, Analyse de Concepts Formels, Règle d'implication multidimensionnelle, Visualisation

Résumé

Pour un producteur agricole, décider d'une pratique impose de considérer celles mises en place afin d'éviter de perturber l'équilibre du système. Il doit donc connaître la diversité des situations culturales. Par exemple, la littérature présente diverses solutions à base de plantes pour contrôler l'infestation d'une culture agricole par une population de bioagresseurs. Choisir une solution qui la repousserait peut la faire migrer vers une culture avoisinante peu attaquée. En comportant plus de 48000 descriptions d'utilisation de plantes à effet pesticide et antibiotique, la base Knomana [Silvie et al., 2021] peut permettre ce choix. Les plateformes logicielles RCAviz [Muller et al. 2022] et RCAvizIR permettent de naviguer dans cette base dont les connaissances ont été classées par l'Analyse de Concepts Relationnels. De façon à représenter fidèlement les données en plusieurs dimensions et faciliter leur interprétation par le producteur agricole, une solution consiste à les exprimer sous forme de règles d'implication multidimensionnelles, une méthode nouvelle issue de l'Analyse de Concepts Formels. Pour une relation ternaire connectant des bioagresseurs, des plantes qui les contrôlent et des cultures protégées, cette méthode permet par exemple d'énoncer les connaissances sous la forme « quand Bioag1 est contrôlé par plant1 sur culture1, alors Bioag1 est également contrôlé par plant2 sur culture1, et par plant3 sur culture2 ». Par exemple : « Quand un Spodoptera frugiperda est contrôlé par Lantana Camara sur arachide, alors il est contrôlé par le neem sur arachide et il est contrôlé par Neem sur maïs. »

Cette méthode se généralise à des relations de dimension supérieure à 3.

L'objectif du stage est de développer une approche et un prototype logiciel de visualisation de connaissances, exprimées sous forme de règles d'implication multidimensionnelles. Ces règles sont produites par un algorithme implémenté en Python. Nous développerons également une stratégie de présentation des règles à l'utilisateur incluant ses centres d'intérêt et d'après la sémantique du contenu des règles.

Profil de la personne recherchée :

Etudiante ou étudiant de Master 2 (informatique ou bioinformatique) ayant des compétences solides en programmation et en analyse de données, ainsi qu'un intérêt pour l'ingénierie des connaissances, l'analyse visuelle (visual analytics) et à trouver des solutions alternatives aux pesticides et antibiotiques de synthèse pour l'agriculture biologique.

Informations complémentaires :

Les plateformes de visualisation RCAviz et RCAvizIR ont été développées par des étudiants de Master 2 grâce à des financements de #Digitag, respectivement en 2020 et 2021. RCAviz (<https://www.hdigitag.fr/fr/rcaviz-visualization-and-exploration-of-conceptual-structures/>) a fait l'objet d'un dépôt logiciel et été communiqué à la conférence CLA en 2022 [Muller et al., 2022]. Le développement de RCAvizIR est en cours de finalisation, puis sera publié. Ces développements complètent le doctorat de P. Keip (2019-2021) également financés par #Digitag [Keip P., 2021]. Ces travaux, ainsi que la présente demande, seront capitalisés dans le cadre du projet SmartFCA (2022-2025, financement ANR) dont un des objectifs est de centraliser et diffuser les approches d'extraction de connaissances basées sur l'analyse formelle de concepts.

Références :

[Silvie et al., 2021] Prototyping a knowledge-based system to identify botanical extracts for plant health in Sub-Saharan Africa. *Plants*, 10 (5):896, 24 p. <https://doi.org/10.3390/plants10050896>

[Muller et al., 2022] RCAviz: Visualizing and Exploring Relational Conceptual Structures. CLA 2022: Sixteenth International Conference on Concept Lattices and Their Applications, Tallinn University of Technology, Jun 2022, Tallin, Estonia. pp.135-148

[Keip, P., 2021] Conversion automatique de modèles et de jeux de données pour l'exploration conceptuelle : Application à une base de connaissances du vivant, Thèse de doctorat, Université de Montpellier.

ENGLISH VERSION

**Navigating the multidimensional implications
of agroecology for animal and plant health decision-making**

Keywords : Agroecology, plant use, Formal Concept Analysis, Multidimensional implication rule, Visualization

Abstract : For a farmer, deciding on a practice requires considering those already in place to avoid disturbing the balance of the system through the introduction of new interactions. He

therefore has to know the diversity of available practices. In the case of crop protection, for example, the literature presents various plant-based solutions (in aqueous or essential oil form) to control the infestation by a pest. For example, choosing a solution that repels a pest population may result in the population moving to a neighboring crop, which is usually not attacked. The Knomana database [Silvie et al., 2021], with over 48,000 descriptions of pesticidal and antibiotic plant use, can enable this choice. The RCAviz [Muller et al., 2022] and RCAvizIR software platforms can be used to navigate through this knowledge base, whose knowledge has been previously classified using Relational Concept Analysis. In order to accurately represent the data of various dimension and to facilitate their interpretation by the farmer, a promising solution is to express them in the form of multidimensional implication rules, a new method derived from Formal Concept Analysis. For a ternary relation (3-D data) relating pests, plants, and protected crop, this method makes it possible, for example, to express knowledge in the form "when Pest1 is controlled by plant1 on crop1, then Pest1 is also controlled by plant2 on crop1, and by plant3 on crop2". This method can be applied to relations of dimension greater than 3. The objective of the internship is to develop a software prototype for visualizing knowledge, expressed in the form of multidimensional implication rules. These rules are produced by an algorithm implemented in Python. We will also develop a strategy so that the rules are presented to the user according to his interests and according to the semantics of the rules' content.

Required profile :

Student in Master studies (computer science or bioinformatics) with strong skills in programming and data analysis, with an interest for knowledge engineering, visual analytics, and to find alternatives to chemical pesticides and antibiotics in organic agriculture.