

Visual Analytics for pesticidal plants usage
Analyse visuelle pour l'usage de plantes à effet pesticide

Offre de stage Master Informatique 2^e année
LIRMM-CIRAD-IRD

stage soutenu par l'Institut Convergences #DigitAg

*Visualisation et Navigation dans des données
spatio-temporelles agro-environnementales
classées par l'analyse de concepts relationnels*

(English version of the scientific subject below)

Contexte : Institut Convergences #DigitAg (programme d'investissement d'avenir)

Le stagiaire est intégré dans la communauté des étudiants de la Graduate School #DigitAg :
- son rapport de stage est publié sur le site internet: <https://www.hdigitag.fr/fr/>
- il est convié à la rencontre annuelle DigitAgora qui se tiendra en 2021 du 31 mai au 2 juin.

Encadrants

M. Huchard, P. Poncelet, A. Sallaberry (LIRMM), Pierre Martin (Cirad), Pierre Silvie (IRD)

Contacts

marianne.huchard@lirmm.fr, pascal.poncelet@lirmm.fr, arnaud.sallaberry@lirmm.fr,
pierre.martin@cirad.fr, pierre.silvie@cirad.fr

Durée 6 mois

Gratification 3600 euros (600 euros par mois), possibilité de financer des missions et du fonctionnement

Profil recherché

Etudiant de Master 2 (informatique) 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.

Description du sujet

Avec l'essor du numérique, la recherche agronomique a produit de nombreux jeux de données sur l'agriculture et l'environnement qu'il s'agit de mobiliser pour développer des outils d'aide à la décision à destination des populations du Nord et du Sud. Parmi ces jeux de données, on peut citer par exemple celui sur les cours d'eau de deux bassins versants français du projet Fresqueau (http://dataqual.engees.unistra.fr/fresqueau_presentation_gb) avec des données spatio-temporelles ou celui sur l'usage des plantes à effet pesticide et antibiotique du projet Knomana (<https://agris.fao.org/agris-search/search.do?recordID=FR2019109314>) pour la santé animale, végétale, humaine et publique dont le modèle de données comporte une structure ternaire.

Pour développer l'outil d'aide à la décision, la méthode de classification utilisée par ces projets pour modéliser la temporalité et la relation ternaire est l'Analyse de Concepts Relationnels (RCA). Par l'emploi de quantificateurs logiques, RCA groupe et classe des groupes d'entités partageant des propriétés et des relations communes, soutenant par exemple le raisonnement par exploration de propriétés et de similarités, le raisonnement par abduction pour la création d'hypothèses, et la recherche de solutions alternatives par voisinage avec des solutions connues. Pour éviter de calculer l'ensemble de la classification pour naviguer et explorer le jeu de données de proche en proche, une méthode de calcul à la volée a été développée (On-demand). Le problème auquel sont confrontées les équipes

porteuses de ces projets, i.e. le LIRMM, l'UPR AIDA, l'UMR IPME et l'ENGEES, est de disposer d'un outil de visualisation et de navigation dans ces données classées par RCA.

Par ailleurs, le LIRMM conduit des recherches en visualisation analytique (Keim et al. 2008). Ce domaine porte sur l'étude des interfaces visuelles interactives permettant d'explorer des jeux de données complexes et hétérogènes afin de faciliter le raisonnement analytique sur les données et ainsi d'en tirer des connaissances (voir par exemple (Accorsi et al. 2014) développé dans le cadre du projet Fresqueau).

Dans le cadre de ce stage, l'objectif est de développer un prototype logiciel de visualisation de jeux de données, comportant des données spatiales et ou temporelles, classés par RCA. Plus précisément, le stagiaire réalisera une visualisation interactive permettant de piloter les calculs à la demande de RCA et d'en afficher, de façon incrémentale, les résultats. Plusieurs approches visuelles seront combinées afin de donner à l'utilisateur une vue d'ensemble de l'espace de connaissances extrait et, à la demande, une vue détaillée de sous-ensembles de la classification calculés à la volée. Différentes méthodes d'interaction (Munzner 2014, chapitres 11-14) et différentes techniques de visualisation de graphes (Tamassia 2013) seront utilisées. Le stagiaire suivra les étapes de conception décrites par Sedlmair et al. 2012 : étude de la littérature, définition du besoin en termes de problème visuel, proposition d'une maquette logicielle, conception du codage visuel et des caractéristiques interactives, développement, déploiement, validation.

Références

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.

Alexandre Bazin, Jessie Carbonnel, Marianne Huchard, Giacomo Kahn, Priscilla Keip, Amirouche Ouzerdine: On-demand Relational Concept Analysis. Formal Concept Analysis - 15th International Conference, ICFCA 2019, Frankfurt, Germany, June 25-28, 2019, Proceedings. LNCS 11511, Springer 2019, ISBN 978-3-030-21461-6: 155-172

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.

Priscilla Keip, Alain Gutierrez, Marianne Huchard, Florence Le Ber, Samira Sarter, Pierre Silvie, Pierre Martin: Effects of Input Data Formalisation in Relational Concept Analysis for a Data Model with a Ternary Relation. Formal Concept Analysis - 15th International Conference, ICFCA 2019, Frankfurt, Germany, June 25-28, 2019, Proceedings. Lecture Notes in Computer Science 11511, Springer 2019, ISBN 978-3-030-21461-6: 191-207

Pierre Martin, Priscilla Keip, Alain Gutierrez, Marianne Huchard, Zakara Ilboudo, et al.: The Knomana knowledge base - A tool to promote exchange of knowledge and identify local plants for addressing sanitary problems in EOA. 5th West Africa Organic Conference (WAOC), Accra, Ghana, 12th -15th November 2019

Tamara Munzner. Visullization Analysis & Design. CRC Press, A K Perters Books, 2014.

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.

Roberto Tamassia, Handbook on Graph Drawing and Visualization. Chapman et Hall / CRC, 2013.

(English version)

Visual Analytics for pesticidal plants usage

Visualisation and Navigation in agro-environmental spatio-temporal data classified by relational concept analysis

With the rise of digital technology, agricultural research has produced numerous datasets on agriculture and on the environment to be mobilized to develop decision-making tools for populations from the North and the South. Among these datasets, there is one on the watercourses of two French watersheds developed by the Fresqueau project (http://dataqual.engees.unistra.fr/fresqueau_presentation_gb) which is spatio-temporal and another one on the uses of plants with pesticidal and antibiotic effect developed by the Knomana project (<https://agris.fao.org/agris-search/search.do?recordID=FR2019109314>) for animal, plant, human and public health whose data model has a ternary structure.

To develop the decision support tool, the classification method used by these projects, to model temporality and ternary relationship, is Relational Concepts Analysis (RCA). Using logical quantifiers, RCA groups and classifies sets of entities sharing common properties and relationships, supporting for example reasoning by exploring properties and similarities, reasoning by abduction to create hypotheses, and the search for alternative solutions by neighborhood with known solutions. To avoid calculating the complete classification to navigate and explore the dataset step by step, an on-demand calculation method has been developed. The problem faced by the team carrying out these projects, i.e. LIRMM, UPR AIDA, UMR IPME and ENGEES, is to have a tool for visualizing and navigating through the data classified by RCA.

Furthermore, the LIRMM conducts research in visual analytics (Keim et al. 2008). This field focuses on the study of interactive visual interfaces enabling the exploration of complex and heterogeneous datasets in order to facilitate analytical reasoning on the data and thus derive knowledge from them (see for example (Accorsi et al. 2014) developed within the Fresqueau project).

The objective of this internship is to develop a software prototype for the visualization of data sets, including spatial and/or temporal data, classified by RCA. More precisely, the trainee will carry out an interactive visualization allowing to pilot the calculations on request of RCA and to display the results in an incremental way. Several visual approaches will be combined in order to give the user an overview of the extracted knowledge space and, as requested by the user, a detailed view of subsets of the classification calculated on the fly. Different interaction methods (Munzner 2014, chapters 11-14) and different graph visualization techniques (Tamassia 2013) will be used. The trainee will follow the design steps described by Sedlmair et al. 2012: i.e. literature review, definition of the need expressed as visual problem, design of visual encodings and interactive features, development, deployment, validation.