

Ingénieur.e de recherche 24 mois : interopérabilité d'outils de fouille de graphes et de données relationnelles

Projet : SmartFCA (ANR), 2022-2026

Laboratoire : IRISA, Rennes

Équipe : LACODAM

Durée : 24 mois, début entre février et avril 2023

Contact : ferre@irisa.fr, peggy.cellier@irisa.fr

Mots-clés : fouille de données, Analyse de concepts formels (FCA), graphes, interopérabilité, données linguistiques

Contexte :

L'Analyse de concepts formels (FCA) [1] est une méthode de découverte de connaissances. Elle est employée dans des tâches d'analyse de données, de fouille de données, de classification ou de recherche d'information ; et appliquée dans domaines variés tels que les sciences de la vie, les sciences humaines ou la linguistique. De multiples extensions de FCA ont été proposées par différentes équipes pour traiter des données complexes telles que des séquences, des trajectoires, des arbres ou des graphes [2]. Au-delà des verrous théoriques et pratiques, il y a un problème d'interopérabilité entre ces différentes extensions, ce qui freine leur adoption et leur composition dans des workflows.

Un objectif important du projet SmartFCA est de rendre ces extensions de FCA interopérables en les encapsulant dans des composants logiciels ayant des interfaces compatibles sur les plans conceptuels et technologiques. Il s'agit aussi d'implémenter une plateforme permettant la construction de workflows à partir des composants. Le partenaire IRISA/Rennes est responsable du composant Graph-FCA [3], une extension de FCA aux données relationnelles et aux graphes. Nous collaborons étroitement avec le partenaire ICube/Strasbourg qui est responsable du composant pour une autre extension de FCA aux données relationnelles, RCA (Relational Concept Analysis) [4].

Un autre objectif du projet est développer des cas d'usage dans divers domaines, pour leur intérêt intrinsèque et pour évaluer la plateforme développée. IRISA/Rennes développera des cas d'usage sur les données linguistiques de langues peu dotées (breton [5,6] et géorgien [7] notamment).

Sujet :

Après une phase de familiarisation avec Graph-FCA et son implémentation actuelle, ainsi que RCA, il s'agira de collaborer avec ICube/Strasbourg pour concevoir une interface compatible entre les deux extensions de FCA (modélisation des entrées/sorties, jeux d'options). Il s'agira ensuite d'encapsuler l'implémentation existante de Graph-FCA dans une API RESTful, en accord avec les normes établies dans le cadre du projet. Il est attendu de la candidate ou du candidat de collaborer avec les autres partenaires du projet dans l'établissement de ces normes, et d'être force de proposition. Il sera également nécessaire de développer des interfaces de test et de démo du composant Graph-FCA pour ne pas dépendre de la plateforme qui ne sera achevée que vers la fin du projet.

Le candidat ou la candidate devra également apporter un support technique et être force de proposition dans les cas d'usage en linguistique (aucune connaissance en linguistique n'est requise). Cela comprend l'aide à la préparation des données, à l'application du composant Graph-FCA et des

autres composants développés dans le projet et à la mise en valeur des résultats, c'est-à-dire des connaissances extraites des données.

Profil du candidat :

Nous cherchons un.e candidat.e motivé.e par une expérience de recherche & développement dans le cadre d'un projet de recherche académique. La formation requise est un doctorat ou un master en informatique.

Expertise requise pour le poste :

- programmation web, surtout backend et Node.js : conception, développement, configuration et documentation
- modèles de données, notamment relationnels et graphes
- outils et méthodes de développement
- travail collaboratif
- rédaction de rapports techniques et présentations orales

Connaissances ou expériences souhaitées :

- extraction de connaissances (fouille de données, data mining, classification)
- programmation Caml ou autre langage fonctionnel (Haskell, Scala, ...)

Qualités attendues : autonomie, rigueur, aptitude à collaborer en présentiel et distanciel avec plusieurs équipes, force de proposition

Bibliographie :

1. Ganter, B., and Wille R. "Formal Concept Analysis – Mathematical Foundations." Springer, 1999.
2. Ferré, S., Huchard, M., Kaytoue, M., Kuznetsov, S. O., & Napoli, A. "Formal concept analysis: from knowledge discovery to knowledge processing." In *A Guided Tour of Artificial Intelligence Research* (pp. 411-445). Springer, Cham, 2020.
3. Ferré, S., and Cellier P. "Graph-FCA: An extension of formal concept analysis to knowledge graphs." *Discrete applied mathematics* 273 (2020): 81-102.
4. Rouane-Hacene, M., Rouane-Hacene, M., Huchard, M., Napoli, A., and Valtchev, P. "Relational concept analysis: mining concept lattices from multi-relational data." *Annals of Mathematics and Artificial Intelligence* 67.1 (2013): 81-108.
5. Foret, A. "Enrichissement de données en breton avec Wordnet." *Celtic Language Technology Workshop within LREC, 2016*.
6. Foret, A., Béchet, D., and Bellynck, V. "Iterated Dependencies in a Breton treebank and implications for a Categorical Dependency Grammar." *Celtic Language Technology Workshop within LREC*. European Language Resources Association, 2022.
7. Ducassé, M. "Kartu-verbs: A semantic web base of inflected georgian verb forms to bypass georgian verb lemmatization issues." *XIX EURALEX conference*. 2021.