



## Sujet de thèse

### Analyse relationnelle de concepts sur un SGBD — Agrégation fondée sur des algèbres qualitatives

**Laboratoire d'accueil :** Laboratoire ICube, UMR 7357 CNRS -- Université de Strasbourg -- INSA -- ENGEES, équipe Science des Données et Connaissances

**Directrice de thèse :** Florence Le Ber (Dr. HDR, ICube)

**Encadrants :** Agnès Braud (Maîtresse de Conférences, ICube), Xavier Dolques (Maître de Conférences, ICube)

**Contact :** envoyer CV, relevés de notes de Master et lettre de motivation à [florence.leber@engees.unistra.fr](mailto:florence.leber@engees.unistra.fr), [agnes.braud@unistra.fr](mailto:agnes.braud@unistra.fr), [dolques@unistra.fr](mailto:dolques@unistra.fr)

**Contexte :** Beaucoup de données à analyser ont un caractère relationnel : données spatiales, temporelles, ou décrivant des liens entre individus. Peu de méthodes sont adaptées à ce type de données, qui nécessitent des approches spécifiques, incluant des techniques d'agrégation. Parmi ces approches, l'analyse relationnelle de concepts est dérivée de l'analyse de concepts formels (ACF) [1], qui est une méthode mathématique de classification, largement appliquée sur différents types de données et dans de nombreux domaines (par exemple [2,3]). Elle consiste, à partir d'une table (appelée contexte) décrivant des objets par des attributs, à construire un treillis de concepts, i.e. des couples (extension ; intension) d'ensembles fermés décrivant les objets et les attributs qui les définissent. L'analyse relationnelle de concepts (ARC) [4] considère deux types de contextes, des contextes objets-attributs et des contextes objets-objets décrivant les relations entre objets. L'ARC étend les contextes objets-attributs par des attributs relationnels de la forme  $qrC$ , où  $q$  est un quantificateur,  $r$  une relation et  $C$  un concept issu du co-domaine de  $r$ . Le résultat de l'ARC est une famille de treillis (un par contexte objets-attributs) reliés entre eux par ces attributs relationnels : un concept d'un treillis représente un groupe d'objets caractérisé par des attributs simples et des attributs relationnels renvoyant à des concepts d'un autre treillis.

**Objectifs de la thèse :** ACF/ARC proposent des outils qui ont fait leurs preuves dans le cadre de différents projets, et produisent des résultats faciles à interpréter par les experts cherchant à analyser des données dans leur domaine. Des améliorations sont toutefois attendues à la fois pour traiter une plus grande volumétrie de données [5] et pour simplifier le processus. Dans ce cadre, la thèse va se focaliser sur les bases de données spatio-temporelles afin de proposer un processus complet d'analyse de ce type de données. Un focus particulier sera fait sur les quantificateurs utilisés dans l'ARC, afin de proposer des quantificateurs spécifiques aux données spatiales et temporelles. On s'appuiera pour cela sur les quantificateurs existants (existential, universel ou de dénombrement [6]) et sur les modèles qualitatifs de l'espace et du temps, telles que les algèbres relationnelles [7].

Habituellement on définit un schéma relationnel pour l'ARC en sélectionnant des tables (extraites de la base de données) définissant des objets et en calculant des relations entre ces objets, ce qui constitue un prétraitement au processus ARC. Dans le cadre de cette thèse, on définira un type de schéma relationnel plus souple, à base de contraintes instanciables, afin de permettre une modification des données traitées sans avoir à recommencer le processus à partir des prétraitements.

Si on prend l'exemple d'une base de données géographiques, où on veut étudier les relations entre parcelles de cultures ayant certaines relations spatiales (*connexe*, *proche*, etc.). Actuellement on extrait les tables objets-attributs (parcelle, culture) et les contextes objets-objets calculés (par exemple les parcelles sont proches si leur distance est inférieure à 500 m). On peut alors créer des attributs relationnels *il-existe-proche.C* en utilisant le quantificateur existentiel. On voudrait pouvoir utiliser un quantificateur spécifique *il-existe-proche* qui

s'applique à un contexte objets-objets plus général comprenant les distances numériques entre parcelles. Ce quantificateur doit pouvoir s'instancier selon différents schémas, où la distance seuil pourrait varier. L'utilisation des algèbres relationnelles permet aussi d'envisager des quantificateurs à différents niveaux de généralité, par exemple il-existe-partie-de peut se spécifier en il-existe-partie-de-tangentielle.

Ce type de quantificateur pourrait aussi s'instancier différemment selon que l'on considère les objets d'un concept ou le concept dans son ensemble, permettant de rendre compte de relations spécifiques entre groupes (comme le font les critères d'agrégation en classification hiérarchique).

Finalement, le travail devrait aboutir à la mise en place d'un processus ARC qui s'applique directement sur une base de données, permettant de traiter plus de données dans un environnement plus convivial pour un analyste. Un autre résultat attendu concerne l'apport théorique sur les quantificateurs spatiaux et temporels et la notion de schéma relationnel instanciable.

#### **Apports attendus :**

- Avancées théoriques sur les quantificateurs et le schéma relationnel
- Développement d'un outil définissant un cadre d'analyse complet
- Expérimentations et validation sur des données réelles (bases de données géographiques et temporelles)

#### **Profil recherché :**

- Master 2 en Informatique ou équivalent
- Formation en logique, représentation de connaissances et programmation
- Curiosité, capacité à appréhender différents domaines et à interagir avec les experts de ces domaines

Pour répondre aux exigences de l'école doctorale de l'Université de Strasbourg, il faut avoir au moins 12 de moyenne en master et licence 3, et être classé dans les 20% premiers de sa promotion de master.

#### **Références :**

- [1] Ganter, B., Wille, R. Formal concept analysis - mathematical foundations. Springer (1999)
- [2] Priss, U. Formal concept analysis in information science. ARIST 40(1), 521–543 (2006)
- [3] [Alam, M., Coulet, A., Napoli, A., Smail-Tabbone, M. Formal Concept Analysis Applied to Transcriptomic Data.](#) CLA 2012, Oct 2012, Fuengirola (Málaga), Spain
- [4] Hacene, M.R., Huchard, M., Napoli, A., Valtchev, P. Relational concept analysis: mining concept lattices from multi-relational data. Ann. Math. Artif. Intell. 67(1), 81–108 (2013)
- [5] Braud, A., Dolques, X., Gutierrez, A., Huchard, M., Keip, P., Le Ber, F., Martin, P., Nica, C., Silvie, P. [Dealing with Large Volumes of Complex Relational Data using RCA.](#) In Complex Data Analytics with Formal Concept Analysis, Rokia Missaoui, Léonard Kwida, Talel Abdesslem (Eds.), Springer (2022)
- [6] Braud, A., Dolques, X., Huchard, M., Le Ber, F. Generalization effect of quantifiers in a classification based on relational concept analysis. Knowledge-Based Systems 160, 119–135 (2018)
- [7] Le Ber, F., Ligozat, G., Papini, O. Raisonnements sur l'espace et le temps : des modèles aux applications, Lavoisier, pp. 419 (2007)