

# Terminaison de la procédure de *Chase* sur les bases de connaissances

Pierre Bourhis, Sylvain Salvati  
Université de Lille, CRIStAL, CNRS, INRIA

## 1 Résumé

Ce stage de M2, s'intéresse aux des algorithmes de *Chase* pour les bases de connaissances. Ces algorithmes jouent un rôle clé pour leur utilisation efficace. Plus précisément, on cherchera à déterminer quand ces algorithmes terminent en particulier lorsque les règles des bases de connaissances sont exprimées dans des logiques gardées.

**Mots clés :** bases de connaissances, *Chase*, logiques gardées

**Équipe :** équipe-projet commune INRIA/CRIStAL Links, Université de Lille1.

## 2 Contexte

Les bases de connaissances permettent de maintenir et de mettre à jour la mémoire technique et les savoirs dans de nombreux domaines. Les connaissances ne se représentent pas simplement par des données, elles sont également constituées des raisonnements ou inférences que les experts font naturellement. Par exemple, une base de connaissances médicales va comporter des descriptions de maladies, des descriptions de symptômes, des descriptions traitements, d'exams, mais également des règles permettant de déduire les complications liées à la co-occurrences de certaines maladies et de leur traitements, sur les risques de certains types de traitements pour les personnes présentant des conditions de santé particulière (enfants, femmes enceintes, diabète, ...), sur les interactions médicamenteuses. . . L'utilisation de ces règles permet non seulement une représentation plus fidèle des connaissances, mais elles présentent également l'avantage technique de limiter la taille des bases de connaissances et de faciliter leur maintenance.

Cependant, lorsque l'on utilise les bases de connaissances, les règles qui leurs sont associées nécessitent de mettre en œuvre des algorithmes de raisonnements automatiques pour répondre aux requêtes qui leurs sont soumises. Afin de bénéficier de l'efficacité des méthodes et des algorithmes liés aux bases de données, il est également possible de transformer la base de connaissances en explicitant toutes les connaissances induites par les règles d'inférence. Cette explicitation transforme la base de connaissances en base de données. Les algorithmes mis en œuvre pour effectuer ce type de transformations portent le nom de *Chase* en anglais.

Malheureusement, il se peut que pour certaines bases de connaissances, cette méthode ne termine pas et génère une base de données infinie. Afin de pouvoir l'appliquer dans des situations concrètes, il s'agit donc de déterminer quand cette méthode termine. Pour cela, on s'intéresse à un type de règles particulières, que l'on appelle gardées. Les logiques gardées peuvent être vues comme des généralisations des logiques modales. Elles sont à la fois suffisamment expressives pour exprimer un grande variété des règles employées par les bases de connaissances et ont de bonnes propriétés algorithmiques.

L'objectif de ce stage consiste à étudier dans quelle mesure l'emploi de règles gardées permet de décider si la méthode de *Chase* termine.

### 3 Problématique

Alors que généralement, le problème de terminaison des algorithmes de *Chase* est indécidable, il a été récemment démontré que ce problème était décidable pour des règles particulières : les TGDs (tuple-generating dependencies) gardées. Ce résultat a été démontré en se basant sur le comportement particulier des algorithmes de *Chase* sur ce type particulier de règles. Cependant, les logiques gardées ont été étudiées comme des généralisations des logiques modales. Elles sont bien plus expressives que ces logiques, mais elles en préservent des caractéristiques essentielles qui leur confèrent généralement de bonnes propriétés algorithmiques. Parmi ces propriétés, on retrouve la propriété de dénombrabilité (dite de Lowenheim-Skolem), ou encore la propriété de modèle arborescent. Cette dernière propriété est particulièrement intéressante pour expliquer l'algorithmique des logiques gardées. En effet, elles permettent de réduire de nombreux problèmes de ces logiques à la décidabilité de la logique monadique du second ordre sur les arbres.

Ainsi, l'objectif du stage est de relier les problèmes de terminaison des algorithmes de *Chase* pour les TGDs gardées à la propriété des modèles arborescents des logiques gardées. Il s'agit de comprendre dans quelle mesure cette propriété joue un rôle dans ce résultat de décidabilité. Si tel est le cas, on cherchera ensuite à généraliser le résultat dans deux directions. Tout d'abord en utilisant une notion plus libérale de garde. Ensuite, nous chercherons à étendre l'algorithme de décision pour des règles plus générales que les TGDs.

### Références

- [1] Michael Benedikt, Balder ten Cate, Thomas Colcombet, and Michael Vanden Boom. The complexity of boundedness for guarded logics. In *LICS*, pages 293–304. IEEE Computer Society, 2015.
- [2] Marco Calautti, Georg Gottlob, and Andreas Pieri. Chase Termination for Guarded Existential Rules. In *PODS'15*, pages 91–103, 2015.
- [3] Tomasz Gogacz and Jerzy Marcinkowski. All-instances termination of chase is undecidable. In *ICALP (2)*, volume 8573 of *Lecture Notes in Computer Science*, pages 293–304. Springer, 2014.
- [4] Tomasz Gogacz and Jerzy Marcinkowski. Termination of oblivious chase is undecidable. *CoRR*, abs/1401.4840, 2014.
- [5] Erich Grädel. Guarded fixed point logics and the monadic theory of countable trees. *Theoretical Computer Science*, 288 :129–152, 2002.