



HAL
open science

Slider : un Raisonneur Incrémental Évolutif

Jules Chevalier, Julien Subercaze, Christophe Gravier, Frederique Laforest

► **To cite this version:**

Jules Chevalier, Julien Subercaze, Christophe Gravier, Frederique Laforest. Slider : un Raisonneur Incrémental Évolutif. 16èmes Journées Francophones Extraction et Gestion des Connaissances, EGC 2016, Jan 2016, Reims, France. hal-01255230

HAL Id: hal-01255230

<https://hal.science/hal-01255230v1>

Submitted on 13 Jan 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Slider : un Raisonneur Incrémental Évolutif

Jules Chevalier*, Julien Subercaze*
Christophe Gravier* Frédérique Laforest*

*Université de Lyon, F-42023, Saint-Étienne, France,
CNRS, UMR5516, Laboratoire Hubert Curien, F-42000, Saint-Étienne, France,
Université de Saint-Étienne, Jean Monnet, F-42000, Saint-Étienne, France.
prenom.nom@univ-st-etienne.fr, <http://laboratoirehubertcurien.fr>

Les solutions existantes pour le raisonnement incrémental souffrent principalement de leur incapacité à prendre en charge des ontologies complexes et ne sont pas conçues pour gérer de grandes quantités de connaissances. Dans cet article, nous présentons Slider (Chevalier et al. (2015)), un raisonneur incrémental évolutif par chaînage avant, permettant de raisonner sur des flux de données sémantiques.

Les principales caractéristiques de Slider sont les suivantes :

Exécution parallèle et passage à l'échelle. Le processus est parallélisé plutôt que distribué, malgré le coût de l'accès concurrent au triplestore.

Limitation des doublons. Les connaissances reçues ou inférées par le raisonneur sont accessibles de manière concurrente par tous les modules du système.

Support de flux de données. L'architecture parallèle de Slider lui permet de recevoir des données de sources statiques et dynamiques.

Indépendance au fragment. Slider supporte nativement les fragments RDFS et ρ df et peut être utilisé pour des fragments plus complexes.

Notre système multi-processus est constitué de modules autonomes, chacun associé à une règle d'inférence, comme le présente la figure 1. Un *triplestore* unique est partagé par ces modules de manière synchronisée. Le partitionnement vertical, introduit par Abadi et al. (2007), permet un accès rapide aux données.

Les triples envoyés au système sont récupérés par le *distributeur général*. Cet élément reçoit les nouveaux triples, les stocke dans le triplestore, et les envoie aux règles pouvant les utiliser. Chaque règle associée à un *buffer* récupère les triples envoyés par le *distributeur général*. Lorsque qu'un *buffer* est plein, ou n'a pas reçu de nouveau triples depuis un temps prédéfini (timeout), un *exécuteur de règle* est instancié afin d'appliquer la règle d'inférence sur les triples présents dans le *buffer*. Ces *buffers* permettent d'améliorer la répartition de charge en appliquant les règles sur des ensembles de triples et non pas pour chaque triple, réduisant ainsi la quantité de processus légers instanciés. Ils assurent également que chaque triple est traité par le système en garantissant, même lors de l'exécution d'une règle, qu'ils sont recueillis par le *buffer*. Les triples inférés par les *exécuteurs de règle* sont ensuite récoltés par le *distributeur* de la règle correspondante. Ce *distributeur* les redistribue aux *buffers* des règles pouvant utiliser ces triples. Afin de déterminer si une règle peut utiliser un triple inféré par une autre

Slider : un Raisonneur Incrémental Évolutif

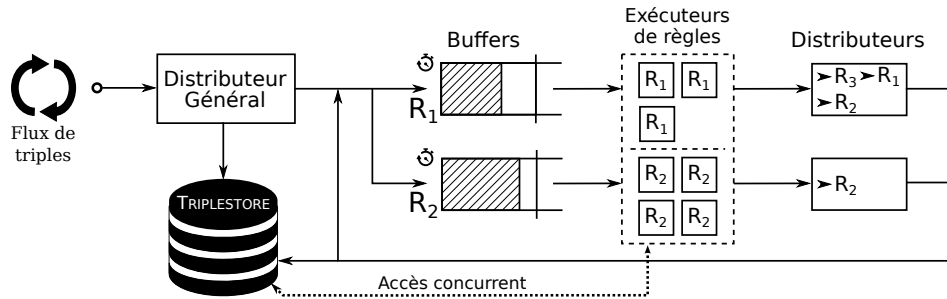


FIG. 1 – Vue d'ensemble de l'architecture du raisonneur Slider

règle, un graphe de dépendance des règles, présenté par Soma et Prasanna (2008), est calculé à l'initialisation du raisonneur.

Nos expérimentations ont montré une amélioration du temps d'inférence de 106,86% sur ρ df et de 36,08% sur RDFS par rapport à la base de comparaison commerciale OWLIM-SE, présentée par Bishop et al. (2011) et grandement utilisée dans les comparatifs de systèmes de raisonnement. Le code source de Slider est disponible librement à l'adresse suivante : [juleschevalier.github.io/slider](https://github.com/juleschevalier/slider). La page d'accueil du projet contient les instructions de reproduction des différentes expérimentations menées.

Références

- Abadi, D. J., A. Marcus, S. R. Madden, et K. Hollenbach (2007). Scalable semantic web data management using vertical partitioning. In *PVLDB*, pp. 411–422.
- Bishop, B., A. Kiryakov, D. Ognyanoff, I. Peikov, Z. Tashev, et R. Velkov (2011). OWLIM : A Family of Scalable Semantic Repositories. *Semantic Web*.
- Chevalier, J., J. Subercaze, C. Gravier, et F. Laforest (2015). Slider : An Efficient Incremental Reasoner. In *Proceedings of SIGMOD*, pp. 1081–1086. ACM.
- Soma, R. et V. K. Prasanna (2008). Parallel inferencing for OWL knowledge bases. In *ICPP'08*, pp. 75–82. IEEE.

Summary

The main drawbacks of current reasoning methods over ontologies are they struggle to provide scalability for large datasets. The batch processing reasoners who provide the best scalability so far are unable to infer knowledge from evolving data. We contribute to solving these problems by introducing Slider, an efficient incremental reasoner. Slider exhibits a performance improvement by more than a 70% compared to the OWLIM-SE reasoner. Slider is conceived to handle expanding data from streams with a growing background knowledge base. It natively supports ρ df and RDFS, and its architecture allows to extend it to more complex fragments with a minimal effort.