Technical Report Number 3

Fast Taxonomic Reasoning Based on Lattice Operations

System Demonstration

Samir Amir and Hassan Aït-Kaci

August 2013

Publication Note

Authors' address:

LIRIS - UFR d'Informatique
Université Claude Bernard Lyon 1
43, boulevard du 11 Novembre 1918
69622 Villeurbanne cedex
France

Email: samir.amir@univ-lyon1.fr
Email: hassan.ait-kaci@univ-lyon1.fr

Phone: +33 (0)4 27 46 57 08

A version of this report appears in the Proceedings for the International Semantic Web Conference's Posters and Demonstrations Track, Sydney, Australia, October 23, 2013 [3].

This work was carried out as part of the $\mathcal{CEDAR}$ Project (Constraint Event-Driven Automated Reasoning) under the *Agence Nationale de la Recherche* (ANR) Chair of Excellence grant N^o **ANR-12-CHEX-0003-01** at the *Université Claude Bernard Lyon 1* (UCBL).

CEDAR **Technical Report Number 3**

Fast Taxonomic Reasoning Based on Lattice Operations

System Demonstration

Samir Amir and Hassan Aït-Kaci

{samir.amir,hassan.ait-kaci}@univ-lyon1.fr

August 2013

Abstract

Taxonomy classification and query answering are the core reasoning services provided by most of the Semantic Web (SW) reasoners. However, the algorithms used by those reasoners are based on Tableau method or Rules. These well-known methods in the literature have already shown their limitations for large-scale reasoning. In this demonstration, we present the part of $\mathcal{CEDAR}$ system that classifies, and reasons about, very large taxonomies using a technique based on lattice operations. This technique makes the $\mathcal{CEDAR}$ reasoner perform on par with the best systems for concept classification, and several orders-of-magnitude more efficiently in terms of response time for query-answering. The experiments were carried out using very large taxonomies (Wikipedia: 111599 sorts, MESH: 286381 sorts, NCBI: 903617 sorts and Biomodels: 182651 sorts).[1] The results achieved by $\mathcal{CEDAR}$ were compared to those obtained by well-known Semantic Web reasoners, namely FaCT++, Pellet, HermiT, TrOWL, SnoRocket and RacerPro.

Keywords: Taxonomic Reasoning; Lattice Operations; Partial-Order Encoding; Semantic Web.

[1] We use "sort" as a synonym for "class" or "atomic concept." In other words, sorts are partially ordered symbols.

Résumé

La classification des taxonomies et les réponses aux requêtes sont les opérations de base sur lesquelles sont fondés la plupart des raisonneurs sémantiques. Cependant, les méthodes utilisées par ces raisonneurs sont basées soit sur la méthode du Tableau, soit sur des régles d'inférence. Ces méthodes, bien connues dans l'état de l'art, ont déjà montré leurs limitations, notamment dans le raisonnement large échelle. Dans cette démonstration, nous présentons le système de $\mathcal{CEDAR}$ pour la classification et le raisonnement à large échelle, basé sur des opérations de treillis algébriques. Les résultats expérimentaux montrent clairement que $\mathcal{CEDAR}$ dépasse de loin les autres raisonneurs en terme de performance. Les expérimentations ont été effectuées en utilisant des taxonomies à large échelle (Wikipedia: 111599 sortes, MESH: 286381 sortes, NCBI: 903617 sortes and Biomodels: 182651 sortes).[2] Les résultats obtenus sont comparés avec ceux de FaCT++, Pellet, HermiT, TrOWL, SnoRocket and RacerPro.

Mots-Clés: Raisonnement taxonomique ; Opérations de Treillis ; Encodage d'ordre partiel ; Web sémantique.

Table of Contents

[2]Nous utilisons le terme "sorte" comme synonyme de "classe" ou "concept atomique."

1 Introduction

This describes a demonstration that illustrates how an implementation of a system based on lattice operations can be used for taxonomic reasoning in a robust and scalable way. Indeed, this challenge was defined in the context of $\mathcal{CEDAR}$ project.[3] This focuses on the part of the $\mathcal{CEDAR}$ system consisting of a Boolean reasoner capable of handling a huge amount of sorts without any noticeable degradation of query evaluation performance. The essential technique we used for implementing the $\mathcal{CEDAR}$ reasoner is based on bit-vector encoding. This method was proposed over 20 years ago for implementing efficient lattice operations [2]. Since the common aspect of all Semantic Web reasoning systems is the representation and processing of taxonomic data, we implemented a taxonomic concept classification and Boolean query-answering system based on the method described above. We made measurements over several very large taxonomies under the exact same conditions for so-called TBox reasoning. A comparative evaluation was conducted to assess the performance of $\mathcal{CEDAR}$ over the mentioned systems which have been implemented by using OWL-API. [4] In terms of query-answering response time, $\mathcal{CEDAR}$ is several orders-of-magnitude more efficient than that of the best existing SW reasoning systems.

2 Lattice Operations for Taxonomic Reasoning

The $\mathcal{CEDAR}$ reasoner is an implementation in Java of the technique described as bottom-up transitive-closure encoding in [1]. This technique consists in representing the elements of a taxonomy (an arbitrary poset) as bit vectors[2]. Thus, each element has a code (a bit vector) carrying a "1" in the position corresponding to the index of any other elements that it subsumes. In this manner, the three Boolean operations on sorts are readily and efficiently performed as their corresponding operations on bit-vectors. This is possible if the bit-vectors encoding the sorts comprising a taxonomy are obtained by computing the reflexive-transitive closure of the "is-a" relation derived from the subsort declarations.

3 Demonstration

This software demonstration shows how $\mathcal{CEDAR}$ differs from existing and known reasoners in terms of classification (Figures 1 and 2) and query answering (Figures 3 and 4) where it is several orders of magnitude more efficient than other reasoners. Developed software integrates six other reasoners to provide a comparison with $\mathcal{CEDAR}$ (HermiT [6], FaCT++ [9], RacerPro [4], TrOWL [8], Pellet [7] and SnoRocket [5] all of which use the OWL-API interface).

The proposed structure of the demonstration is as follows.

[3] Constraint Event-Driven Automated Reasoning—http://cedar.liris.cnrs.fr

[4] http://owlapi.sourceforge.net

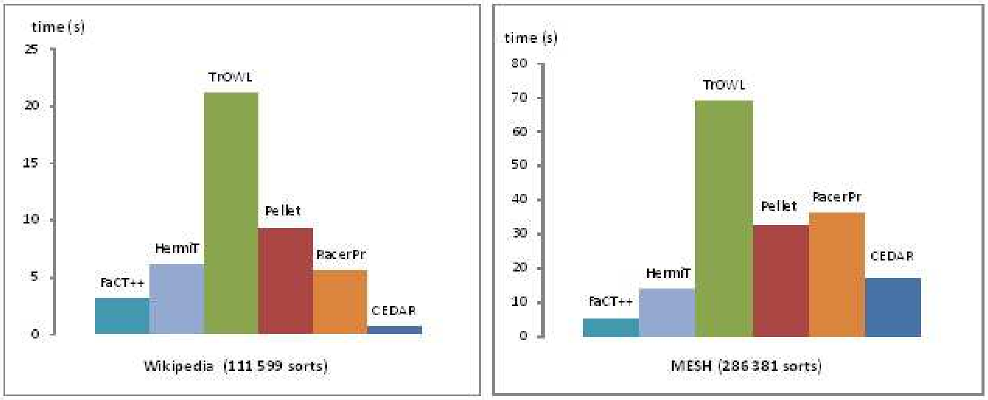

Figure 1: Classification time per reasoner for the “Wikipedia” and “MeSH” taxonomies

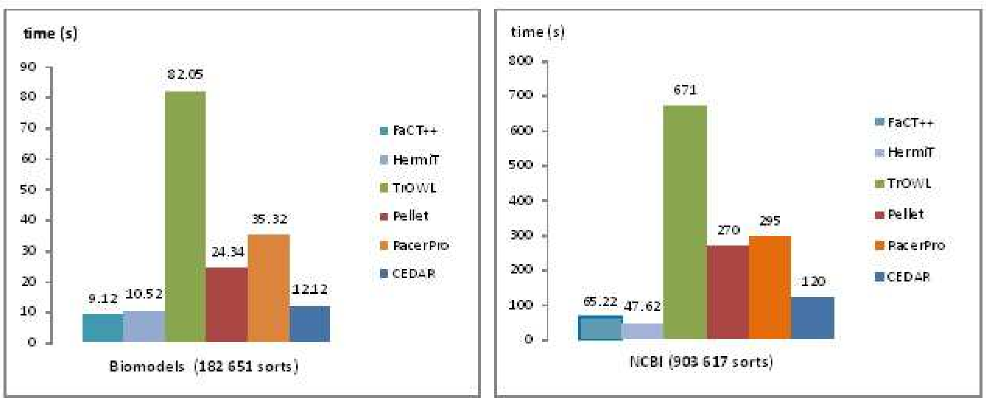

Figure 2: Classification time per reasoner for the “Biomodels” and “NCBI” taxonomies

- Classification performance using very large taxonomies as Wikipedia[5] (111599 sorts), NCBI[6](903617 sorts), MESH[7] (286381 sorts) and Biomodels[8] (182651 sorts). The demonstration shows the results illustrated in Figures 1 and 2 where $\mathcal{CEDAR}$ is always among the best three out of six reasoners.
- Query Answering using boolean queries (`and`, `or` and `not`) involving a large number of concepts (up to 100 concepts). The obtained results can be be compared with those of traditional reasoners as shown in Figures 3 and 4.
- With $\mathcal{CEDAR}$, there is no need to perform a classification each time. A classified taxonomy can be saved and reused.

[5] http://www.h-its.org/english/research/nlp/download/wikitaxonomy.php
[6] http://www.ncbi.nlm.nih.gov/Taxonomy/taxonomyhome.html/
[7] http://www.nlm.nih.gov/mesh/meshhome.html
[8] https://code.google.com/p/sbmlharvester/

- Detecting Cycles—the demonstration also shows how to detect cycles in taxonomies, which are a particular case of inconsistency resulting from modeling errors.

A web service is available at http://cedar.univ-lyon1.fr for anyone wishing to verify our claims.

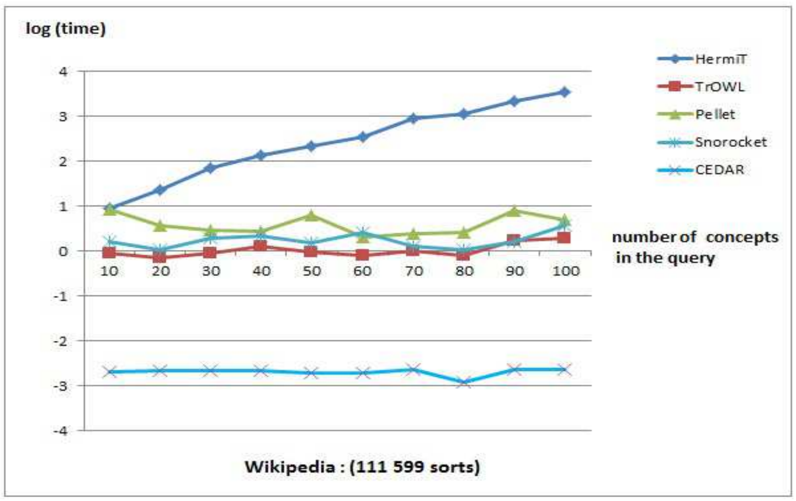

Figure 3: Query response time per reasoner for the "Wikipedia" taxonomy

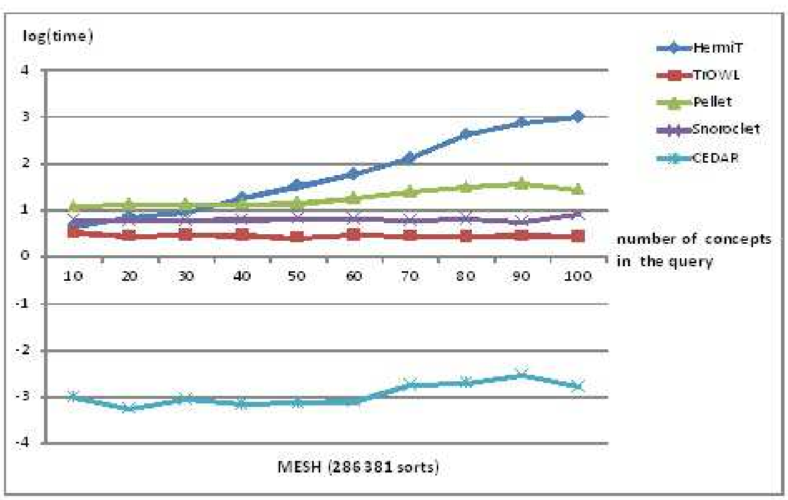

Figure 4: Query response time per reasoner for the "MeSH" taxonomy

References

[1] Hassan Aït-Kaci and Samir Amir. Classifying and querying very large taxonomies—a comparative study to the best of our knowledge. *CEDAR* Technical Report Number 2, Laboratoire d'Informatique en Image et Systèmes d'Information, Université Claude Bernard Lyon 1, Villeurbanne, France, May 2013. [Available online[9]].

[2] Hassan Aït-Kaci, Robert Boyer, Patrick Lincoln, and Roger Nasr. Efficient implementation of lattice operations. *ACM Transactions on Programming Languages and Systems*, 11(1):115–146, January 1989. [Available online[10]].

[3] Samir Amir and Hassan Aït-Kaci. *CEDAR*: a fast taxonomic reasoner based on lattice operations. In Eva Blomqvist and Tudor Groza, editors, *Proceedings of the 12th International Semantic Web Conference Posters and Demonstrations Track*, pages 9–12, Sydney, Australia, October 23, 2013. CEUR Workshop Proceedings, Volume 1035. [Available online[11]].

[4] Volker Haarslev, Kay Hidde, Ralf Möller, and Michael Wessel. The RacerPro knowledge representation and reasoning system. *Semantic Web Journal*, 1:1–11, March 2011. [Available online[12]].

[5] Michael J. Lawley and Cyril Bousquet. Fast classification in Protégé: Snorocket as an OWL 2 EL reasoner. In Thomas Meyer, Mehmet A. Orgun, and Kerry Taylor, editors, *Proceedings of the 2nd Australasian Ontology Workshop: Advances in Ontologies*, pages 45–50, Adelaide, Australia, December 2010. AOW'10, ACS. [Available online[13]].

[6] Rob Shearer, Boris Motik, and Ian Horrocks. HermiT: A highly-efficient OWL reasoner. In Ulrike Sattler and Cathy Dolbear, editors, *Proceedings of the 5th International Workshop on OWL Experiences and Directions*, Karlsruhe, Germany, October 2008. OWLED'08, CEUR Workshop Proceedings. [Available online[14]].

[7] Evren Sirin, Bijan Parsia, Bernardo Cuenca Grau, Aditya Kalyanpur, and Yarden Katz. Pellet: A practical OWL-DL reasoner. *Journal of Web Semantics*, 5(2):51–53, June 2007. This is a summary; full paper: [Available online[15]].

[8] Edward Thomas, Jeff Z. Pan, and Yuan Ren. TrOWL: Tractable OWL 2 reasoning infrastructure. In Lora Aroyo, Grigoris Antoniou, Eero Hyvnen, Annette ten Teije, Heiner Stuckenschmidt, Liliana Cabral, and Tania Tudorache, editors, *Proceedings of the 7th Extended Semantic Web Conference*, pages 431–435, Heraklion, Greece, May-June 2010. ESWC'10, Springer-Verlag. [Available online[16]].

[9] Dmitry Tsarkov and Ian Horrocks. FaCT++ description logic reasoner: System description. In Ulrich Furbach and Natarajan Shankar, editors, *Proceedings of the 3rd International Joint conference on Automated Reasoning*, pages 292–297, Seattle, WA, USA, August 2006. IJCAR'06, Springer-Verlag. [Available online[17]].

Acknowledgements: The authors wish to thank Mohand-Saïd Hacid for his comments.

[9] http://cedar.liris.cnrs.fr/papers/ctr2.pdf
[10] http://www.hassan-ait-kaci.net/pdf/encoding-toplas-89.pdf
[11] http://ceur-ws.org/Vol-1035/
[12] http://www.semantic-web-journal.net/sites/default/files/swj109_3.pdf
[13] http://krr.meraka.org.za/~aow2010/Lawley-etal.pdf
[14] http://www.cs.ox.ac.uk/ian.horrocks/Publications/download/2008/ShMH08b.pdf
[15] http://pellet.owldl.com/papers/sirin05pellet.pdf
[16] http://homepages.abdn.ac.uk/jeff.z.pan/pages/pub/TPR2010.pdf
[17] http://www.cs.ox.ac.uk/Ian.Horrocks/Publications/download/2006/TsHo06a.pdf

Technical Report Number 3

Fast Taxonomic Reasoning
Based on Lattice Operations

Samir Amir and Hassan Aït-Kaci

August 2013