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].

Copyright © 2013 by the CEDAR Project.

This work was carried out as part of the CEDAR Project (Constraint Event-Driven Automated Reasoning) under the Agence Nationale de la Recherche (ANR) Chair of Excellence grant No ANR-12-CHEX-0003-01 at the Université Claude Bernard Lyon 1 (UCBL). It may not be copied nor reproduced in whole or in part for any commercial purpose. Permission to copy in whole or in part without payment of fee is granted for non-profit educational and research purposes provided that all such whole or partial copies include the following: a notice that such copying is by permission of the UCBL, with an acknowledgement of the authors and individual contributors to the work; and all applicable portions of the copyright notice. Copying, reproducing, or republishing for any other purpose shall require a license with payment of a fee to the UCBL. All rights reserved.
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 CEDAR system that classifies, and reasons about, very large taxonomies using a technique based on lattice operations. This technique makes the 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). The results achieved by 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.

---

1We use “sort” as a synonym for “class” or “atomic concept.” In other words, sorts are partially ordered symbols.
Résumé


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

Table of Contents

1 Introduction 1
2 Lattice Operations for Taxonomic Reasoning 1
3 Demonstration 1

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 CEDAR project. This focuses on the part of the 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 CEDAR reasoner is based on bit-vector encoding. This method was proposed over 20 years ago for implementing efficient lattice operations. 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 CEDAR over the mentioned systems which have been implemented by using OWL-API. In terms of query-answering response time, CEDAR is several orders-of-magnitude more efficient than that of the best existing SW reasoning systems.

2 Lattice Operations for Taxonomic Reasoning

The 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. 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 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 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.

---

2http://owlapi.sourceforge.net

---
• Classification performance using very large taxonomies as Wikipedia (111599 sorts), NCBI (903617 sorts), MESH (286381 sorts) and Biomodels (182651 sorts). The demonstration shows the results illustrated in Figures 1 and 2 where 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 compared with those of traditional reasoners as shown in Figures 3 and 4.

• With CEDAR, there is no need to perform a classification each time. A classified taxonomy can be saved and reused.
- 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—


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

9http://cedar.liris.cnrs.fr/papers/ctr2.pdf
11http://ceur-ws.org/Vol-1035/
14http://www.cs.ox.ac.uk/ian.horrocks/Publications/download/2008/ShMH08b.pdf
Technical Report Number 3
Fast Taxonomic Reasoning
Based on Lattice Operations
Samir Amir and Hassan Aït-Kaci
August 2013