Artificial Intelligence 215 (2014) 120–140

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

(Non-)Succinctness of uniform interpolants of general terminologies in the description logic $\mathcal{EL}$ ☆

Nadeschda Nikitina [a,*], Sebastian Rudolph [b,1]

[a] University of Oxford, Department of Computer Science, Wolfson Building, Parks Road, Oxford, OX1 3QD, UK
[b] Technische Universität Dresden, Fakultät Informatik, Nöthnitzer Straße 46, Dresden, 01062, Germany

ARTICLE INFO

Article history:
Received 15 December 2012
Received in revised form 14 June 2014
Accepted 17 June 2014
Available online 25 June 2014

Keywords:
Ontologies
Knowledge representation
Automated reasoning
Description logics
Uniform interpolation
Forgetting
$\mathcal{EL}$

ABSTRACT

$\mathcal{EL}$ is a popular description logic, used as a core formalism in large existing knowledge bases. Uniform interpolants of knowledge bases are of high interest, e.g. in scenarios where a knowledge base is supposed to be partially reused. However, to the best of our knowledge no procedure has yet been proposed that computes uniform $\mathcal{EL}$ interpolants of general $\mathcal{EL}$ terminologies. Up to now, also the bound on the size of uniform $\mathcal{EL}$ interpolants has remained unknown. In this article, we propose an approach to computing a finite uniform interpolant for a general $\mathcal{EL}$ terminology if it exists. To this end, we develop a quadratic representation of $\mathcal{EL}$ TBoxes as regular tree grammars. Further, we show that, if a finite uniform $\mathcal{EL}$ interpolant exists, then there exists one that is at most triple exponential in the size of the original TBox, and that, in the worst case, no smaller interpolants exist, thereby establishing tight worst-case bounds on their size. Beyond showing these bounds, the notions and results established in this paper also provide useful insights for designing efficient ontology reformulation algorithms, for instance, within the context of module extraction.

1. Introduction

With the wide-spread adoption of ontological modelling by means of the W3C-specified OWL Web Ontology Language [2], description logics (DLs, [3,4]) have developed into one of the most popular family of formalisms employed for knowledge representation and reasoning [5–8]. For application scenarios where scalability of reasoning is of utmost importance, specific tractable sublanguages (the so-called *profiles* [9]) of OWL have been put into place, among them OWL 2 EL which in turn is based on DLs of the $\mathcal{EL}$ family [10,11].

In view of the practical deployment of OWL and its profiles [12–14], non-standard reasoning services for supporting modelling activities gain in importance. An example of such reasoning services supporting knowledge engineers in different tasks is that of *uniform interpolation*: given a theory using a certain vocabulary, and a subset of "relevant terms" of that vocabulary, find a theory (referred to as a *uniform interpolant*, short: UI) that uses only the relevant terms and gives rise to the same consequences (expressible via relevant terms) as the original theory. Intuitively, this provides a view on the ontology where all irrelevant (asserted as well as implied) statements have been filtered out.

☆ This paper is an invited revision of a paper which first appeared at the ECAI 2012 conference [1].

* Corresponding author. Tel.: +44 (1865) 610747; fax: +44 (1865) 283531.
E-mail addresses: nadeschda.nikitina@cs.ox.ac.uk (N. Nikitina), sebastian.rudolph@tu-dresden.de (S. Rudolph).
[1] Tel.: +49 (351) 463 38516; fax: +49 (351) 463 32827.

http://dx.doi.org/10.1016/j.artint.2014.06.005
0004-3702/

Uniform interpolation has many applications within ontology engineering. For instance, it can help ontology engineers understand existing ontological specifications by visualizing implicit dependencies between relevant concepts and roles, as used, for instance, for interactive ontology revision [15]. In particular for understanding and developing complex knowledge bases, e.g., those consisting of *general concept inclusions* (GCIs), appropriate tool support of this kind would be beneficial. Another application of uniform interpolation is ontology reuse: given an ontology that is to be reused in a different scenario, most likely not all aspects of this ontology are relevant to the new usage requirements. In combination with module extraction, uniform interpolation can be used to reduce the amount of irrelevant information within an ontology employed in a new context.

For DL-Lite, the problem of uniform interpolation has been investigated [16,17] and a tight exponential bound on the size of uniform interpolants has been shown. Lutz and Wolter [18] propose an approach to uniform interpolation in expressive description logics such as $\mathcal{ALC}$ featuring general terminologies showing a tight triple-exponential bound on the size of uniform interpolants. Koopmann and Schmidt [19] and Ludwig and Konev [20] propose practical approaches to computing uniform interpolants in expressive description logics. For the lightweight description logic $\mathcal{EL}$, the problem of uniform interpolation has, however, not been solved. To the best of our knowledge, the only existing approach [21] to uniform interpolation in $\mathcal{EL}$ is restricted to terminologies containing each concept symbol at most once on the left-hand side of concept inclusions and additionally satisfying particular acyclicity conditions which are sufficient, but not necessary for the existence of a uniform interpolant. Recently, Lutz, Seylan and Wolter [22] proposed an ExpTime procedure for deciding, whether a finite uniform $\mathcal{EL}$ interpolant exists for a particular general terminology and a particular set of relevant terms. However, the authors do not address the actual computation of such a uniform interpolant. Up to now, also the bounds on the size of uniform $\mathcal{EL}$ interpolants have remained unknown.

In this paper, we propose a worst-case-optimal approach to computing a finite uniform $\mathcal{EL}$ interpolant for a general terminology. Our approach is based on proof theory and regular tree languages. We develop a grammar representation of $\mathcal{EL}$ TBoxes. These grammars are quadratic in the size of the initial TBox and capture all of its logical consequences except for a certain kind of *weak consequences* – consequences that can be trivially derived from other logical consequences but are not equivalent to those. We show via a proof-theoretic analysis that the tree languages generated by the proposed grammars indeed capture all non-weak consequences of the initial terminology expressed using the set of relevant terms.

Further, we show that certain finite subsets of the languages generated by these grammars can be transformed into a uniform $\mathcal{EL}$ interpolant of at most triple exponential size, if such a finite uniform $\mathcal{EL}$ interpolant exists for the given terminology and a set of terms. We also show that, in the worst-case, no shorter interpolants exist, thereby establishing tight bounds on the size of uniform interpolants in $\mathcal{EL}$.

It should be noted that the notions and results presented in this article go beyond the mere purpose of showing the triple exponential blowup and have practical applications. In fact, the proposed grammars have served as a basis for a module extraction tool in follow-up work by Nikitina and Glimm [23]. Within this tool, the insights gained in the present article are taken into account to derive a blowup-avoiding algorithm for a kind of partial uniform interpolation that conditionally eliminates concept symbols one by one after a careful analysis.

The article is structured as follows: In Section 2, we recall the necessary preliminaries on $\mathcal{EL}$. In Section 3, we introduce a calculus for deriving general subsumptions in $\mathcal{EL}$ terminologies, which is used as a major tool in the proofs of this work. Section 4 formally introduces the notion of inseparability and defines the task of uniform interpolation. Section 5 demonstrates that the smallest uniform interpolants in $\mathcal{EL}$ can be triple exponential in the size of the original knowledge base. In Section 6.1, we describe a normalisation of terminologies that enables a representation of non-weak logical consequences as languages of regular tree grammars. In Section 6.2, we recall the necessary preliminaries on regular tree languages/grammars and introduce regular tree grammars representing subsumees and subsumers of concept symbols, which are the basis for computing uniform $\mathcal{EL}$ interpolants as shown in Section 6.3. In the same section, we also show the upper bound on the size of uniform interpolants. After giving an overview of related work in Section 7, we summarize the contributions in Section 8 and discuss some ideas for future work. This is a revised and extended version of our previous paper [1] and contains technical enhancements, a more detailed argumentation, examples and the full proofs.

2. Preliminaries

In this section, we formally introduce the description logic $\mathcal{EL}$, and recall some of its well-known properties. Let N_C and N_R be countably infinite and mutually disjoint sets called *concept symbols* and *role symbols*, respectively. $\mathcal{EL}$ *concepts* C are defined by

$$C ::= A \mid C \sqcap C \mid \exists r.C$$

where A and r range over $N_C \cup \{\top\}$ and N_R, respectively. In the following, C, D, E, F and G can denote arbitrary concepts, while A, B can only denote concept symbols (i.e., concepts from N_C) or $\top$. We use the term *simple concept* to refer to a simpler form of $\mathcal{EL}$ concepts defined by $C_s ::= A \mid \exists r.A$, where A and r range over $N_C \cup \{\top\}$ and N_R, respectively.