# On the (Non-)Succinctness of Uniform Interpolation in General $\mathcal{EL}$ Terminologies

Nadeschda Nikitina, Sebastian Rudolph

Karlsruhe Institute of Technology
Karlsruhe, Germany
{nikitina,rudolph}@kit.edu

## 1 Introduction

In view of the practical deployment of OWL [9] based on description logics [2], the importance of non-standard reasoning services for supporting ontology engineers was pointed out, for instance, in [8]. An example of such reasoning services is that of *uniform interpolation*: given a theory using a certain vocabulary, and a subset $\Sigma$ of "relevant terms" of that vocabulary, find a theory that uses only $\Sigma$ terms and gives rise to the same consequences (expressible via $\Sigma$) as the original theory. In particular for the understanding and the development of complex knowledge bases, e.g., those consisting of *general concept inclusions* (GCIs), the appropriate tool support would be beneficial. We consider the task of uniform interpolation in the very lightweight description logic $\mathcal{EL}$, the basic member of the $\mathcal{EL}$ family [1] which provides the logical backbone of the OWL EL profile. The existing related approaches ([3, 6, 4]) do not provide a solution for the task of uniform interpolation in general $\mathcal{EL}$ terminologies. Up to now, also the bounds on the size of uniform $\mathcal{EL}$ interpolants have been unknown. We propose a worst-case-optimal approach to computing a finite uniform $\mathcal{EL}$ interpolant for a general terminology. After a normalization, we construct two regular tree grammars generating subsumees and subsumers of atomic concepts interpreted as tree languages. Using a Gentzen-style proof calculus for general subsumptions in $\mathcal{EL}$, we show that, in case a uniform interpolant exists, the corresponding sublanguages with an exponential bound on the role depth are sufficient to obtain a uniform $\mathcal{EL}$ interpolant of at most triple exponential size. Further, we show that, in the worst-case, no smaller interpolants exist, thereby establishing the triple exponential tight bounds on the size of uniform interpolants in $\mathcal{EL}$. This is a report on our recent work accepted at ECAI 2012 [7].

## 2 Preliminaries

Let $N_C$ and $N_R$ be countably infinite and mutually disjoint sets of concept symbols and role symbols. An $\mathcal{EL}$ concept $C$ is defined as $C ::= A|\top|C \sqcap C|\exists r.C$, where $A$ and $r$ range over $N_C$ and $N_R$, respectively. In the following, we use symbols $A, B$ to denote atomic concepts and $C, D$ to denote arbitrary concepts. A *terminology* or *TBox* consists of *concept inclusion* axioms $C \sqsubseteq D$ and *concept equivalence* axioms $C \equiv D$ used as a shorthand for $C \sqsubseteq D$ and $D \sqsubseteq C$. While knowledge bases in general can also include a specification of individuals with the corresponding concept and role assertions

(ABox), in this paper we do not consider them. The signature of an $\mathcal{EL}$ concept $C$ or an axiom $\alpha$, denoted by $\mathrm{sig}(C)$ or $\mathrm{sig}(\alpha)$, respectively, is the set of concept and role symbols occurring in it. To distinguish between the set of concept symbols and the set of role symbols, we use $\mathrm{sig}_C(C)$ and $\mathrm{sig}_R(C)$, respectively. The signature of a TBox $\mathcal{T}$, in symbols $\mathrm{sig}(\mathcal{T})$ (correspondingly, $\mathrm{sig}_C(\mathcal{T})$ and $\mathrm{sig}_R(\mathcal{T})$), is defined analogously. The semantics of the above introduced DL constructs is standard and can be found, for instance, in [2].

In this paper, we investigate uniform interpolation based on concept- inseparability, i.e., the aim is to preserve all $\Sigma$-concept inclusions. Thus, the task of *uniform interpolation* is defined as follows: Given a signature $\Sigma$ and a TBox $\mathcal{T}$, determine a TBox $\mathcal{T}'$ with $\mathrm{sig}(\mathcal{T}') \subseteq \Sigma$ such that for all $\mathcal{EL}$ concepts $C, D$ with $\mathrm{sig}(C) \cup \mathrm{sig}(D) \subseteq \Sigma$ holds: $\mathcal{T} \models C \sqsubseteq D$ iff $\mathcal{T}' \models C \sqsubseteq D$. $\mathcal{T}'$ is also called a *uniform $\mathcal{EL}$ $\Sigma$-interpolant* of $\mathcal{T}$. In practice, uniform interpolants are required to be finite, i.e., expressible by a finite set of finite axioms using only the language constructs of $\mathcal{EL}$.

## 3  Lower Bound

While deciding the existence of uniform interpolants in $\mathcal{EL}$ is exponential [4], i.e., one exponential less complex than the same decision problem for the more complex logic $\mathcal{ALC}$ [6], the size of uniform interpolants remains triple-exponential. We demonstrate that this is in fact the lower bound by the means of the following example (obtained by a slight modification of an example given in [5] originally demonstrating a double exponential lower bound in the context of conservative extensions).

*Example 1.* The $\mathcal{EL}$ TBox $\mathcal{T}_n$ for a natural number $n$ is given by

$$A_1 \sqsubseteq \overline{X_0} \sqcap ... \sqcap \overline{X_{n-1}} \tag{1}$$
$$A_2 \sqsubseteq \overline{X_0} \sqcap ... \sqcap \overline{X_{n-1}} \tag{2}$$
$$\sqcap_{\sigma \in \{r,s\}} \exists \sigma.(\overline{X_i} \sqcap X_0 \sqcap ... \sqcap X_{i-1}) \sqsubseteq X_i \qquad i < n \tag{3}$$
$$\sqcap_{\sigma \in \{r,s\}} \exists \sigma.(X_i \sqcap X_0 \sqcap ... \sqcap X_{i-1}) \sqsubseteq \overline{X_i} \qquad i < n \tag{4}$$
$$\sqcap_{\sigma \in \{r,s\}} \exists \sigma.(\overline{X_i} \sqcap \overline{X_j}) \sqsubseteq \overline{X_i} \quad j < i < n \tag{5}$$
$$\sqcap_{\sigma \in \{r,s\}} \exists \sigma.(X_i \sqcap \overline{X_j}) \sqsubseteq X_i \quad j < i < n \tag{6}$$
$$X_0 \sqcap ... \sqcap X_{n-1} \sqsubseteq B \tag{7}$$

In $\mathcal{T}_n$, the atomic concepts $X_i$ and $\overline{X_i}$ represent the bit number $i$ of a binary counter being set and unset, respectively. Axiom 3 ensures that an unset bit will be set in the successor number, if all smaller bits are already set. The subsequent Axiom 4 ensures that a set bit will be unset in the successor number, if all smaller bits are also set. Axioms 5 and 6 ensure that in all other cases, bits do not flip. For instance, Axiom 5 states that, if any bit before bit $i$ is still unset, then bit $i$ will remain unset also in the successor number. If we now consider sets $C_i$ of concept descriptions inductively defined by $C_0 = \{A_1, A_2\}$, $C_{i+1} = \{\exists r.C_1 \sqcap \exists s.C_2 \mid C_1, C_2 \in C_i\}$, then we find that $|C_{i+1}| = |C_i|^2$ and consequently $|C_i| = 2^{(2^i)}$. Thus, the set $C_{2^n-1}$ contains triply exponentially many different concepts, each of which is doubly exponential in the size of $\mathcal{T}_n$ (intuitively, we obtain concepts having the shape of binary trees of exponential depth, thus having doubly exponentially many leaves, each of which can be endowed with $A_1$ or $A_2$, giving rise to

triply exponentially many different such trees). It can be shown that for each concept $C \in C_{2^n-1}$ it holds $\mathcal{T}_n \models C \sqsubseteq B$ and that there cannot be a smaller uniform interpolant for $\mathcal{T}_n$ w.r.t. the signature $\Sigma = \{A_1, A_2, B, r, s\}$ than $\{C \sqsubseteq B \mid C \in C_{2^n-1}\}$.

Hence we have found a class $\mathcal{T}_n$ of TBoxes giving rise to uniform interpolants of triple-exponential size in terms of the original TBox. In the following, we show that this is also an upper bound by providing a procedure for computing uniform interpolants with a triple-exponentially bounded output.

## 4 Upper Bound

The upper bound can be shown by providing an algorithm, which computes a uniform interpolant, in case it exists, of at most triple-exponential size in the size of the original TBox. The algorithm relies on a normalization, which assigns to each sub-expression occurring in the original TBox and not being equivalent to any atomic concept a fresh concept name. This can be done recursively by replacing sub-expressions $C_1 \sqcap ... \sqcap C_n$ and $\exists r.C$ by fresh concept symbols until each axiom in the TBox $\mathcal{T}$ is one of $\{A \sqsubseteq B, A \equiv B_1 \sqcap ... \sqcap B_n, A \equiv \exists r.B\}$, where $A, B, B_i \in \text{sig}_C(\mathcal{T}) \cup \{\top\}$ and $r \in \text{sig}_R(\mathcal{T})$. Given a normalized TBox additionally extended with classification results, we can show using a deduction calculus for $\mathcal{EL}$ terminologies that the uniform interpolant UI can be obtained from the sets of subsumers and subsumees of all atomic concepts in $\mathcal{T}$ as follows.

**Definition 1.** *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox and , for each $A \in \text{sig}_C(\mathcal{T})$, let $R_1(A)$ and $R_2(A)$ be the set of subsumees and the set of subsumers of $A$ in $\mathcal{T}$. Then, the $\mathcal{EL}$ TBox $\text{UI}(\mathcal{T}, \Sigma, R_1, R_2)$ is given by*

$\{C \sqsubseteq A \mid A \in \Sigma, C \in R_1(A)\} \cup \{A \sqsubseteq D \mid A \in \Sigma, D \in R_2(A)\} \cup$
$\{C \sqsubseteq D \mid \text{there is } A \notin \Sigma, C \in R_1(A), D \in R_2(A)\}.$

In our approach, we represent the (possibly infinite) sets of subsumees and subsumers as tree languages $L(G)$ generated by regular tree grammars $G$, where concept expressions $C$ are interpreted as a trees according to their term structure.

**Theorem 1.** *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox, $\Sigma$ a signature. For each $A \in \text{sig}_C(\mathcal{T})$, we can compute from $\mathcal{T}$ in exponential time a grammar $G^{\exists}(\mathcal{T}, \Sigma, A)$ generating subsumees of A and a grammar $G^{\sqsubseteq}(\mathcal{T}, \Sigma, A)$ generating subsumers of A with the following properties:*

- *$G^{\exists}(\mathcal{T}, \Sigma, A)$ and $G^{\sqsubseteq}(\mathcal{T}, \Sigma, A)$ are exponentially bounded in the size of $\mathcal{T}$, while the number of non-terminals corresponds to the number of atomic concepts in $\mathcal{T}$.*
- *For each C with $\text{sig}(C) \subseteq \Sigma$ such that $\mathcal{T} \models C \sqsubseteq A$ there is a concept $C'$ generated by $G^{\exists}(\mathcal{T}, \Sigma, A)$ such that C can be obtained from $C'$ by adding arbitrary conjuncts to arbitrary sub-expressions.*
- *Each D satisfying $\text{sig}(D) \subseteq \Sigma$ and $\mathcal{T} \models A \sqsubseteq D$ is generated by $G^{\sqsubseteq}(\mathcal{T}, \Sigma, A)$.*

While the languages generated by the grammars are usually infinite, we require finite subsets of $L(G^{\exists}(\mathcal{T}, \Sigma, A))$ and $L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$ to obtain the corresponding upper bound. Based on the following lemma presented in [4], we obtain a bound on the role depth of minimal uniform $\mathcal{EL}$ interpolants, allowing us to restrict the role depth of relevant elements in $L(G^{\exists}(\mathcal{T}, \Sigma, A))$ and $L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$:

**Lemma 1.** *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox, $\Sigma$ a signature. There exists a uniform $\mathcal{EL}$ $\Sigma$-interpolant of $\mathcal{T}$ if and only if there exists a uniform $\mathcal{EL}$ $\Sigma$-interpolant $\mathcal{T}'$ of $\mathcal{T}$ whose maximal role depth is exponentially bounded by $|\mathcal{T}|$.*

Based on this bound and the size of $G^{\sqsupseteq}(\mathcal{T},\Sigma,A),G^{\sqsubseteq}(\mathcal{T},\Sigma,A)$, we can describe a way to materialize a role-depth-bounded part of $G^{\sqsupseteq}(\mathcal{T},\Sigma,A),G^{\sqsubseteq}(\mathcal{T},\Sigma,A)$ into subsumer and subsumee sets $R_1(A)$ and $R_2(A)$, respectively, obtaining the following result:

**Theorem 2.** *Given an $\mathcal{EL}$ TBox $\mathcal{T}$ and a signature $\Sigma$, there exists a uniform $\mathcal{EL}$ $\Sigma$-interpolant of $\mathcal{T}$ iff there exists a uniform $\mathcal{EL}$ $\Sigma$-interpolant $\mathcal{T}'$ with $|\mathcal{T}'| \in O(2^{2^{2^{|\mathcal{T}|}}})$.*

## 5  Summary

In this paper, we summarize an approach to computing uniform interpolants of general $\mathcal{EL}$ terminologies based on proof theory and regular tree languages. Moreover, we noted that, if a finite uniform $\mathcal{EL}$ interpolant exists, then there exists one of at most triple exponential size in terms of the original TBox, and that, in the worst-case, no shorter interpolant exists, thereby establishing the triple exponential tight bounds.

## References

1. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI-05). pp. 364–369 (2005)
2. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, second edn. (2007)
3. Konev, B., Walther, D., Wolter, F.: Forgetting and uniform interpolation in large-scale description logic terminologies. In: Proc. of the 21st Int.Joint Conf. on Artificial Intelligence (IJCAI-09). pp. 830–835 (2009)
4. Lutz, C., Seylan, I., Wolter, F.: An automata-theoretic approach to uniform interpolation and approximation in the description logic EL. In: Proc. of the 13th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR-12) (2012)
5. Lutz, C., Wolter, F.: Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. Journal of Symbolic Computation 45(2), 194–228 (2010)
6. Lutz, C., Wolter, F.: Foundations for uniform interpolation and forgetting in expressive description logics. In: Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI-11). pp. 989–995 (2011)
7. Nikitina, N., Rudolph, S.: ExpExpExplosion: Uniform interpolation in general EL terminologies. In: Proc. of the 20th European Conf. on Artificial Intelligence (ECAI-12) (2012), to appear. Extended version available via http://dl.dropbox.com/u/10637748/11.pdf
8. Nikitina, N., Rudolph, S., Glimm, B.: Reasoning-supported interactive revision of knowledge bases. In: Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI-11). pp. 1027–1032 (2011)
9. OWL Working Group, W.: OWL 2 Web Ontology Language: Document Overview. W3C Recommendation (27 October 2009), available at `http://www.w3.org/TR/owl2-overview/`