ECAI 2012
Luc De Raedt et al. (Eds.)

doi:10.3233/978-1-61499-098-7-618

ExpExpExplosion: Uniform Interpolation in General $\mathcal{EL}$ Terminologies

Nadeschda Nikitina and **Sebastian Rudolph**[1]

Abstract. Although $\mathcal{EL}$ is a popular logic used in large existing knowledge bases, 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 bounds on the size of uniform $\mathcal{EL}$ interpolants remain unknown. In this paper, we propose an approach based on proof theory and the theory of formal tree languages to computing a finite uniform interpolant for a general $\mathcal{EL}$ terminology if it exists. Further, we show that, if such 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 shorter interpolants exist, thereby establishing the triple exponential tight bounds on their size.

1 Introduction

With the wide-spread adoption of ontological modeling by means of the W3C-specified OWL Web Ontology Language [16], description logics [2, 17] have developed into one of the most popular family of formalisms employed for knowledge representation and reasoning.

For application scenarios where scalability of reasoning is of utmost importance, specific tractable sublanguages (the so-called *profiles* [12]) of OWL have been put into place, among them OWL EL which in turn is based on DLs of the $\mathcal{EL}$ family [3, 1].

In view of this practical deployment of OWL and its profiles, the importance of non-standard reasoning services for supporting knowledge engineers in modeling a particular domain or in understanding existing models by visualizing implicit dependencies between concepts and roles was pointed out by the research community [4, 15]. An example of such reasoning services supporting knowledge engineers in different activities is that of *uniform interpolation*: given a theory using a certain vocabulary, and a subset of "relevant terms" of that vocabulary, find a theory that uses only the relevant terms and gives rise to the same consequences (expressible via relevant terms) 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.

In our paper, we consider the task of uniform interpolation in the very lightweight description logic $\mathcal{EL}$. An existing approach [7] to uniform interpolation in $\mathcal{EL}$ is restricted to terminologies containing each atomic concept at most once on the left-hand side of concept inclusions and additionally satisfying sufficient, but not necessary acyclicity conditions. Lutz and Wolter [11] propose an approach to uniform interpolation in expressive description logics such as $\mathcal{ALC}$ featuring general terminologies, which, however does not solve the problem of uniform interpolation in $\mathcal{EL}$. Recently, Lutz, Seylan and Wolter [9] 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 remain unknown.

In this paper, we propose a worst-case-optimal approach based on proof theory and the theory of formal tree languages to computing a finite uniform $\mathcal{EL}$ interpolant for a general terminology. For this purpose, we introduce regular tree grammars representing subsumees and subsumers of atomic concepts, which, after a sequence of nonterminal replacements, 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. Further, by the means of an example we show that, in the worst-case, no shorter interpolants exist, thereby establishing the triple exponential tight bounds on the size of uniform interpolants in $\mathcal{EL}$.

The paper is structured as follows: In Section 2, we recall the necessary preliminaries on $\mathcal{EL}$ and regular tree languages/grammars. Section 3 formally introduces the notion of inseparability, defines the task of uniform interpolation and provides an example that demonstrates that the smallest uniform interpolants in $\mathcal{EL}$ can be triple exponential in the size of the original knowledge base. In Section 5, we introduce regular tree grammars representing subsumees and subsumers of atomic concepts, which are the basis for computing uniform $\mathcal{EL}$ interpolants as shown in Section 6. In the same section, we also show the upper bound on the size of uniform interpolants. We summarize the contributions in Section 7 and discuss some ideas for future work. Detailed proofs are available in the extended version of this paper [14].

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 abstract from ABoxes and concentrate on TBoxes. The signature of an $\mathcal{EL}$ concept C or an axiom α, denoted by $\mathrm{sig}(C)$ or $\mathrm{sig}(\alpha)$,

[1] Karlsruhe Institute of Technology, Karlsruhe, Germany, email: {nikitina,rudolph}@kit.edu

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. Next, we recall the semantics of the above introduced DL constructs, which is defined by the means of interpretations. An interpretation $\mathcal{I}$ is given by the domain $\Delta^\mathcal{I}$ and a function $\cdot^\mathcal{I}$ assigning each concept $A \in N_C$ a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$ and each role $r \in N_R$ a subset $r^\mathcal{I}$ of $\Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The interpretation of $\top$ is fixed to $\Delta^\mathcal{I}$. The interpretation of an arbitrary $\mathcal{EL}$ concept is defined inductively, i.e., $(C \sqcap D)^\mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I}$ and $(\exists r.C)^\mathcal{I} = \{x \mid (x, y) \in r^\mathcal{I}, y \in C^\mathcal{I}\}$. An interpretation $\mathcal{I}$ satisfies an axiom $C \sqsubseteq D$ if $C^\mathcal{I} \subseteq D^\mathcal{I}$. $\mathcal{I}$ is a model of a TBox, if it satisfies all of its axioms. We say that a TBox $\mathcal{T}$ entails an axiom α (in symbols, $\mathcal{T} \models \alpha$), if α is satisfied by all models of $\mathcal{T}$.

Tree Languages and Regular Tree Grammars

A ranked alphabet is a pair ($\mathcal{F}$, Arity) where $\mathcal{F}$ is a finite set and Arity is a mapping from $\mathcal{F}$ into $\mathbb{N}$. $T(\mathcal{F})$ denotes the set of ground terms over the alphabet $\mathcal{F}$. Let $\mathcal{X}_n$ be a set of n variables. A term $C \in T(\mathcal{F}, \mathcal{X}_n)$ containing each variable from $\mathcal{X}_n$ at most once is called a context. We denote by $C(\mathcal{F})$ the set of contexts containing a single variable. A regular tree grammar $G = (S, \mathcal{N}, \mathcal{F}, R)$ is composed of a start symbol S, a set $\mathcal{N}$ of non-terminal symbols (non-terminal symbols have arity 0) with $S \in \mathcal{N}$, a ranked alphabet $\mathcal{F}$ of terminal symbols with a fixed arity such that $\mathcal{F} \cap \mathcal{N} = \emptyset$, and a set R of derivation rules of the form $X \to \beta$ where β is a tree of $T(\mathcal{F} \cup \mathcal{N})$ and $X \in \mathcal{N}$. Given a regular tree grammar $G = (S, \mathcal{N}, \mathcal{F}, R)$, the derivation relation $\to_G$ associated to G is a relation on pairs of terms of $T(\mathcal{F} \cup \mathcal{N})$ such that $s \to_G t$ if and only if there is a rule $X \to \alpha \in R$ and there is a context C such that $s = C[X]$ and $t = C[\alpha]$. The language generated by G, denoted by $L(G)$ is a subset of $T(\mathcal{F})$ which can be reached by successive derivations starting from the start symbol, i.e. $L(G) = \{s \in T \mid S \to^+ s\}$ with $\to^+$ the transitive closure of $\to$. We write $\to$ instead of $\to_G$ when the grammar G is clear from the context. For further details, we refer the reader, for instance, to [5].

3 Uniform Interpolation

Formally, the term uniform interpolation is defined based on the notion of *inseparability*. Two TBoxes, $\mathcal{T}_1$ and $\mathcal{T}_2$, are inseparable w.r.t. a signature Σ if they have the same Σ-consequences, i.e., consequences whose signature is a subset of Σ. Depending on the particular application requirements, the expressivity of those Σ consequences can vary from subsumption queries and instance queries to conjunctive queries. In this paper, we investigate uniform interpolation based on concept inseparability of general $\mathcal{EL}$ terminologies defined analogously to previous work on inseparability, e.g., [8] or [7], as follows:

Definition 1 *Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two general $\mathcal{EL}$ TBoxes and Σ a signature. $\mathcal{T}_1$ and $\mathcal{T}_2$ are concept-inseparable w.r.t. Σ, in symbols $\mathcal{T}_1 \equiv^c_\Sigma \mathcal{T}_2$, if for all $\mathcal{EL}$ concepts C, D with $sig(C) \cup sig(D) \subseteq \Sigma$ holds $\mathcal{T}_1 \models C \sqsubseteq D$, iff $\mathcal{T}_2 \models C \sqsubseteq D$.*

Given a signature Σ and a TBox $\mathcal{T}$, the aim of uniform interpolation is to determine a TBox $\mathcal{T}'$ with $\mathrm{sig}(\mathcal{T}') \subseteq \Sigma$ such that $\mathcal{T} \equiv^c_\Sigma \mathcal{T}'$. $\mathcal{T}'$ is also called a *uniform $\mathcal{EL}$ Σ-interpolant* of $\mathcal{T}$. In practise, uniform interpolants are required to be finite, i.e., expressible by a finite set of finite axioms using only the language constructs of $\mathcal{EL}$. As demonstrated by the following example, in the presence of cyclic concept inclusions, a finite uniform $\mathcal{EL}$ Σ-interpolant might not exist for a particular TBox $\mathcal{T}$ and a particular Σ.

Example 1 *Consider uniform interpolants of the TBox $\mathcal{T} = \{A' \sqsubseteq A, A \sqsubseteq A'', A \sqsubseteq \exists r.A, \exists s.A \sqsubseteq A\}$. w.r.t. $\Sigma = \{s, r, A', A''\}$. We obtain an infinite chain of consequences $A' \sqsubseteq \exists r.\exists r.\exists r....A''$ and $\exists s.\exists s.\exists s....A' \sqsubseteq A''$ containing nested existential quantifiers of unbounded depth.*

It is interesting that, while deciding the existence of uniform interpolants in $\mathcal{EL}$ [9] is one exponential less complex than the same decision problem for the more complex logic $\mathcal{ALC}$ [11], the size of uniform interpolants remains triple-exponential due to the unavailability of disjunction. We demonstrate that this is in fact the lower bound by the means of the following example (obtained by a slight modification of the corresponding example given in [10] originally demonstrating a double exponential lower bound in the context of conservative extensions).

Example 2 *The $\mathcal{EL}$ TBox $\mathcal{T}_n$ for a natural number n is given by*

$$A_1 \sqsubseteq \overline{X_0} \sqcap ... \sqcap \overline{X_{n-1}} \tag{1}$$

$$A_2 \sqsubseteq \overline{X_0} \sqcap ... \sqcap \overline{X_{n-1}} \tag{2}$$

$$\sqcap_{\sigma \in \{r,s\}} \exists \sigma.(\overline{X_i} \sqcap X_0 \sqcap ... \sqcap X_{i-1}) \sqsubseteq X_i \qquad i < n \tag{3}$$

$$\sqcap_{\sigma \in \{r,s\}} \exists \sigma.(X_i \sqcap X_0 \sqcap ... \sqcap X_{i-1}) \sqsubseteq \overline{X_i} \qquad i < n \tag{4}$$

$$\sqcap_{\sigma \in \{r,s\}} \exists \sigma.(\overline{X_i} \sqcap \overline{X_j}) \sqsubseteq \overline{X_i} \quad j < i < n \tag{5}$$

$$\sqcap_{\sigma \in \{r,s\}} \exists \sigma.(X_i \sqcap \overline{X_j}) \sqsubseteq X_i \quad j < i < n \tag{6}$$

$$X_0 \sqcap ... \sqcap X_{n-1} \sqsubseteq B \tag{7}$$

If we now consider sets $\mathcal{C}_i$ of concept descriptions inductively defined by $\mathcal{C}_0 = \{A_1, A_2\}$,$\mathcal{C}_{i+1} = \{\exists r.C_1 \sqcap \exists s.C_2 \mid C_1, C_2 \in \mathcal{C}_i\}$, then we find that $|\mathcal{C}_{i+1}| = |\mathcal{C}_i|^2$ and consequently $|\mathcal{C}_i| = 2^{(2^i)}$. Thus, the set $\mathcal{C}_{2^n-1}$ 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, which gives rise to triply exponentially many different such trees). Then it can be shown that for each concept $C \in \mathcal{C}_{2^n-1}$ holds $\mathcal{T}_n \models C \sqsubseteq B$ and that there cannot be a smaller uniform interpolant w.r.t. the signature $\Sigma = \{A_1, A_2, B, r, s\}$ than the one containing all these GCIs (for a proof, see [14]).

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 Normalization

Similarly to other proof-theoretic approaches [1, 6, 7], we will make use of normalizations that restrict the syntactic form of TBoxes. We decompose complex axioms into syntactically simpler ones. The decomposition is realized 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 $A, B, B_i \in \mathrm{sig}_C(\mathcal{T}) \cup \{\top\}$ and $r \in \mathrm{sig}_R(\mathcal{T})$. For this purpose, we introduce a minimal required set of fresh concept symbols

N_D and the corresponding definition axioms $\{A' \equiv C' \mid A' \in N_D\}$ for each $A' \in N_D$ and the corresponding concept C' replaced by A'.

In what follows, we assume that knowledge bases are normalized and refer to $\text{sig}_C(\mathcal{T}) \cup N_D$ as $\text{sig}_C(\mathcal{T})$. Since concept symbols in N_D are fresh, they do not appear in Σ. W.l.o.g., in what follows we assume that $\mathcal{EL}$ concepts do not contain any equivalent concepts in conjunctions and that equivalent concept symbols have been replaced by a single representative of the corresponding equivalence class. The following lemma postulates the close semantic relation between a TBox and its normalization.

Lemma 1 *Any $\mathcal{EL}$ TBox $\mathcal{T}$ can be extended into a normalized TBox $\mathcal{T}'$ such that each model of $\mathcal{T}'$ is a model of $\mathcal{T}$ and each model of $\mathcal{T}$ can be extended into a model of $\mathcal{T}'$.*

Proof Sketch. All concepts in N_D are defined, i.e., their meaning is uniquely determined by the meaning of subconcepts (concepts that occur in $\mathcal{T}$) of the original TBox $\mathcal{T}$. □

The following lemma motivates the usefulness of the normalization for the computation of uniform interpolants. In particular, it allows us to restrict the information necessary for the uniform interpolation to the sets of subsumers and subsumees of all atomic concepts in the TBox.

Lemma 2 *Let $\mathcal{T}$ be normalized $\mathcal{EL}$ TBox and C, D two $\mathcal{EL}$ concepts with $sig(C) \cup sig(D) \subseteq sig(\mathcal{T})$ such that $\mathcal{T} \models C \sqsubseteq D$. For any $A \in sig_C(\mathcal{T})$, let $\mathtt{Pre}(A) = \{M \subseteq sig_C(\mathcal{T}) \mid \mathcal{T} \models \bigsqcap_{B_i \in M} B_i \sqsubseteq A\}$. W.l.o.g., assume that*

$$C = \bigsqcap_{1 \le j \le n} A_j \sqcap \bigsqcap_{1 \le k \le m} \exists r_k.E_k$$

for $A_j \in sig_C(\mathcal{T})$ and $r_k \in sig_R(\mathcal{T})$, E_k $\mathcal{EL}$ concepts with $sig(E_k) \subseteq sig(\mathcal{T})$ for $1 \le k \le m$. For all conjuncts D_i of D, the following is true: If $D_i \in sig_C(\mathcal{T})$, there is a set $M \in \mathtt{Pre}(D_i)$ of atomic concepts such that for each element B of M holds at least one of the conditions [A1]-[A2]:

(A1) There is an A_j in C such that $A_j = B$.

(A2) There are r_k, E_k and there exists $B' \in sig_C(\mathcal{T})$ such that $\mathcal{T} \models E_k \sqsubseteq B'$ and $B \equiv \exists r_k.B' \in \mathcal{T}$.

If $D_i = \exists r'.D'$ for $r' \in sig_R(\mathcal{T})$ and D' an $\mathcal{EL}$ concept , at least one of the conditions [A3]-[A4] holds:

(A3) There are r_k, E_k such that $r_k = r'$ and $\mathcal{T} \models E_k \sqsubseteq D'$.

(A4) There is a $B \in sig_C(\mathcal{T})$ such that $\mathcal{T} \models B \sqsubseteq \exists r'.D'$ and $\mathcal{T} \models C \sqsubseteq B$.

Proof. The proof is based on a Gentzen-style calculus for $\mathcal{EL}$ complete for subsumptions between arbitrary $\mathcal{EL}$ concepts shown in Fig. 1. We consider all rules, that could have been the last rule applied in order to derive the above sequent and show the lemma by induction on the length of the proof. □

Lemma 2 allows us, on the one hand, to prove the completeness of grammars introduced in the next section, and, on the other hand, to show that the TBox computed in Section 6 by combining subsumees and subsumers into subsumption axioms indeed entails all Σ-consequences of $\mathcal{T}$.

5 Grammar Representation of Subsumees and Subsumers

In order to obtain a finite uniform interpolant from the infinite sets of subsumees and subsumers, a finite representation for these sets is required. In this section, we show how, for a signature Σ, the sets of Σ-subsumees and Σ-subsumers of each atomic concept in a normalized $\mathcal{EL}$ TBox $\mathcal{T}$ can be described as languages generated by regular tree grammars on ranked unordered trees with finite sets of derivation rules later on transformed into a finite uniform interpolant. For the definition of the grammars, we uniquely represent each atomic concept $A \in \text{sig}_C(\mathcal{T})$ by a non-terminal $\mathfrak{n}_A$ (and denote the set of all non-terminals by $\mathcal{N}^{\mathcal{T}} = \{\mathfrak{n}_x | x \in \text{sig}_C(\mathcal{T}) \cup \{\top\}\}$). In what follows, we use the ranked alphabet $\mathcal{F} = (\text{sig}_C(\mathcal{T}) \cap \Sigma) \cup \{\top\} \cup \{\exists r \mid r \in \text{sig}_R(\mathcal{T}) \cap \Sigma\} \cup \{\sqcap_i \mid i \le n\}$, where atomic concepts in $\text{sig}_C(\mathcal{T}) \cap \Sigma$ are constants, $\exists r$ for $r \in \text{sig}_R(\mathcal{T}) \cap \Sigma$ are unary functions and $\sqcap_i$ are functions of the arity i bounded by $n = |\text{sig}_C(\mathcal{T})| \cdot (|\text{sig}_R(\mathcal{T})| + 1)$, i.e., the number of all possible simple concepts in $\mathcal{T}$ (atomic concepts and all existential restrictions on atomic concepts). The restriction to the maximum arity of n is w.l.o.g., since we can always split longer conjunctions into a nested conjunction with at most n elements in each sub-expression. In the following, it will be convenient to simply write $\sqcap$ if the arity of the corresponding function is clear from the context. Clearly, every $\mathcal{EL}$ concept C with $\text{sig}(C) \subseteq \Sigma$ and at most n conjuncts in each sub-expression has a unique representation by the means of the above functions. We denote such a term representation of C using $\mathcal{F}$ by t_C.

In what follows, we use a substituting function $\sigma_{\mathcal{T},\mathcal{F}} : \{C \mid \text{sig}(C) \subseteq \text{sig}(\mathcal{T})\} \to T(\mathcal{F}, \mathcal{N}^{\mathcal{T}})$ by $\sigma_{\mathcal{T},\mathcal{F}}(C) = t_C\{\mathfrak{n}_\top/\top, \mathfrak{n}_{B_1}/B_1, ..., \mathfrak{n}_{B_n}/B_n\}$, where $B_1, ..., B_n$ are all atomic sub-expressions of C. Note that $\sigma_{\mathcal{T},\mathcal{F}}$ is injective, therefore, its inverse is also a function. If the TBox and the set of non-terminals are clear from the context, we will denote such a representation of a concept C simply by $\sigma(C)$, and its inverse by $\sigma^-(t)$ for $t \in T(\mathcal{F}, \mathcal{N}^{\mathcal{T}})$. In the following we will assume $\sigma^-(t)$ to be extended to partially ground terms and ground terms.

Since concepts are represented as terms, we extend the generated languages by associative variants of concept expressions. For this purpose, in addition to the TBox axioms and classification results, we include in our grammars the subsumees and subsumers of each atomic concept having the form of simple conjunctions, i.e., conjunctions of simple concepts. As we will see in the next section, to obtain a uniform interpolant and derive the corresponding upper bound, in the case of subsumees, it is sufficient to capture all associative variants of subsumees not being obtained by adding arbitrary conjuncts to arbitrary sub-expressions (rule ANDL in Fig. 1). In fact, in general, adding arbitrary conjuncts to arbitrary sub-expressions allows us to obtain subsumees being conjunctions of unbounded size, which would cause the corresponding language to contain terms with $\sqcap$-

$$\frac{}{C \sqsubseteq C}(\text{Ax}) \quad \frac{}{C \sqsubseteq \top}(\text{AxTop})$$

$$\frac{D \sqsubseteq E}{C \sqcap D \sqsubseteq E}(\text{AndL})$$

$$\frac{C \sqsubseteq E \quad C \sqsubseteq D}{C \sqsubseteq D \sqcap E}(\text{AndR})$$

$$\frac{C \sqsubseteq D}{\exists r.C \sqsubseteq \exists r.D}(\text{Ex})$$

$$\frac{C \sqsubseteq E \quad E \sqsubseteq D}{C \sqsubseteq D}(\text{Cut})$$

Figure 1. Gentzen-style proof system for general $\mathcal{EL}$ terminologies.

functions of unbounded arity and make the definition of the grammar unnecessary complex. Therefore, we do not include such subsumees into our grammars. For this reason, it is sufficient in the case of subsumees to consider conjunctions of atomic concepts only, denoted by $\mathtt{Pre}(A) = \{M \subseteq \mathrm{sig}_C(\mathcal{T}) \mid \mathcal{T} \models \bigsqcap_{B_i \in M} B_i \sqsubseteq A\}$.

In contrast to that, to be able to derive the upper bound, we have to include all subsumers into our grammars. Since weakening of subsumers (see rule ANDR in Fig. 1) does not require $\sqcap$-functions of unbounded arity, this can be done by the means of a minor extension: in addition to conjunctions of atomic concepts, we take into account existential restrictions with atomic concepts, formed from the elements of the set $\mathtt{Post}_{\mathrm{Base}}(A) = \{A' \in \mathrm{sig}_C(\mathcal{T}) \cup \{\top\} \mid \mathcal{T} \models A \sqsubseteq A'\} \cup \{\exists r.A' \mid A' \in \mathrm{sig}_C(\mathcal{T}) \cup \{\top\}, \mathcal{T} \models A \sqsubseteq \exists r.A', r \in \Sigma\}$ and $\mathtt{Post}(A) = 2^{\mathtt{Post}_{\mathrm{Base}}(A)}$. Thereby, we obtain the following definition.

Definition 2 *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox, Σ a signature. Further, let $\mathtt{Pre}(A) = \{M \subseteq sig_C(\mathcal{T}) \mid \mathcal{T} \models \bigsqcap_{B_i \in M} B_i \sqsubseteq A\}$, $\mathtt{Post}_{Base}(A) = \{A' \in sig_C(\mathcal{T}) \cup \{\top\} \mid \mathcal{T} \models A \sqsubseteq A'\} \cup \{\exists r.A' \mid A' \in sig_C(\mathcal{T}) \cup \{\top\}, \mathcal{T} \models A \sqsubseteq \exists r.A', r \in \Sigma\}$ and $\mathtt{Post}(A) = 2^{\mathtt{Post}_{Base}(A)}$. Further, for each $B \in sig_C(\mathcal{T})$, let $R^{\sqsupseteq}$ be given by*

(GL1) $\mathfrak{n}_B \rightarrow B$ *if* $B \in \Sigma$,
(GL2) $\mathfrak{n}_B \rightarrow \mathfrak{n}_{B'}$ *for all* $\{B'\} \in \mathtt{Pre}(B)$,
(GL3) $\mathfrak{n}_B \rightarrow \sqcap(\mathfrak{n}_{B'_1}, ..., \mathfrak{n}_{B'_n})$ *for all* $\{B'_1, ..., B'_n\} \in \mathtt{Pre}(B)$ *with* $n \geq 1$,
(GL4) $\mathfrak{n}_B \rightarrow \exists r(\mathfrak{n}_{B'})$ *for all* B' *with* $B \equiv \exists r.B' \in \mathcal{T}$ *and* $r \in sig_R(\mathcal{T}) \cap \Sigma$.

Let $R^{\sqsubseteq}$ be given for all $B \in sig_C(\mathcal{T}) \cup \{\top\}$ by

(GR1) $\mathfrak{n}_B \rightarrow B$ *if* $B \in \Sigma \cup \{\top\}$,
(GR2) $\mathfrak{n}_B \rightarrow \sigma(C)$ *for all* $\{C\} \in \mathtt{Post}(B)$,
(GL3) $\mathfrak{n}_B \rightarrow \sqcap(\mathfrak{n}_{C'_1}, ..., \mathfrak{n}_{C'_n})$ *for all* $\{C'_1, ..., C'_n\} \in \mathtt{Post}(B)$ *with* $n \geq 1$.

For each $A \in sig_C(\mathcal{T})$, the regular tree grammar $G^{\sqsupseteq}(\mathcal{T}, \Sigma, A)$ is then given by $(\mathfrak{n}_A, \mathcal{N}^{\mathcal{T}}, \mathcal{F}, R^{\sqsupseteq})$, and the regular tree grammar $G^{\sqsubseteq}(\mathcal{T}, \Sigma, A)$ is given by $(\mathfrak{n}_A, \mathcal{N}^{\mathcal{T}}, \mathcal{F}, R^{\sqsubseteq})$.

We denote the set of tree grammars $\{G^{\sqsupseteq}(\mathcal{T}, \Sigma, A) \mid A \in \mathrm{sig}_C(\mathcal{T})\}$ by $\mathbb{G}^{\sqsupseteq}(\mathcal{T}, \Sigma)$ and the set $\{G^{\sqsubseteq}(\mathcal{T}, \Sigma, A) \mid A \in \mathrm{sig}_C(\mathcal{T})\}$ by $\mathbb{G}^{\sqsubseteq}(\mathcal{T}, \Sigma)$.

Since $\mathrm{sig}(\mathcal{T})$ is finite, all elements of `Pre` and `Post` can be effectively computed. For the construction of grammars the following result holds.

Theorem 1 *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox, Σ a signature. $\mathbb{G}^{\sqsupseteq}(\mathcal{T}, \Sigma)$ and $\mathbb{G}^{\sqsubseteq}(\mathcal{T}, \Sigma)$ can be computed from $\mathcal{T}$ in exponential time and are exponentially bounded in the size of $\mathcal{T}$.*

Proof Sketch. The exponentially bounded size and time hold basically due to the exponential number of elements in `Pre` and `Post` and tractable reasoning in $\mathcal{EL}$[1]. $\square$

The following example demonstrates the grammar construction.

Example 3 *For $\mathcal{T}$ and Σ from Example 1, we obtain a normalized TBox $\mathcal{T}' = \{A' \sqsubseteq A, A \sqsubseteq A'', A \sqsubseteq B, B \equiv \exists r.A, B' \equiv \exists s.A, B' \sqsubseteq A\}$, which yields $\mathtt{Pre} = \{(A, \{A', B'\}), (A'', \{A', B', A\}), (A', \{\}), (B, \{A', A\}), (B', \{\})\}$, $\mathtt{Post}_{Base} = \{(A, \{A'', B, \top, \exists r(\mathfrak{n}_A), \exists r(\mathfrak{n}_\top)\}), (A', \{A, A'', B, \top, \exists r(\mathfrak{n}_A), \exists r(\mathfrak{n}_\top)\}), (B, \{\top, \exists r(\mathfrak{n}_A), \exists r(\mathfrak{n}_\top)\}), (A'', \{\top\}), (B', \{A'', A, \top, \exists s(\mathfrak{n}_A), \exists s(\mathfrak{n}_\top)\})\}$ and the following set of transitions for $R^{\sqsupseteq}$:*

$$\mathfrak{n}_B \rightarrow \mathfrak{n}_A$$

$$\begin{array}{rl}
\mathfrak{n}_{A''} \rightarrow \mathfrak{n}_{A'} & \mathfrak{n}_A \rightarrow \mathfrak{n}_{B'} \\
\mathfrak{n}_{A''} \rightarrow \mathfrak{n}_A & \mathfrak{n}_A \rightarrow \mathfrak{n}_{A'} \\
\mathfrak{n}_{A''} \rightarrow \mathfrak{n}_{B'} & \mathfrak{n}_B \rightarrow \mathfrak{n}_{A'} \\
\mathfrak{n}_{A''} \rightarrow A'' & \mathfrak{n}_{A'} \rightarrow A' \\
\mathfrak{n}_{B'} \rightarrow \exists s(\mathfrak{n}_A) & \mathfrak{n}_B \rightarrow \exists r(\mathfrak{n}_A) \\
\mathfrak{n}_A \rightarrow \sqcap(\mathfrak{n}_{A'}, \mathfrak{n}_{B'}) & \mathfrak{n}_B \rightarrow \sqcap(\mathfrak{n}_{A'}, \mathfrak{n}_A) \\
\mathfrak{n}_{A''} \rightarrow \sqcap(\mathfrak{n}_{A'}, \mathfrak{n}_A) & \mathfrak{n}_{A''} \rightarrow \sqcap(\mathfrak{n}_A, \mathfrak{n}_{B'}) \\
\mathfrak{n}_{A''} \rightarrow \sqcap(\mathfrak{n}_{A'}, \mathfrak{n}_{B'}) & \mathfrak{n}_{A''} \rightarrow \sqcap(\mathfrak{n}_A, \mathfrak{n}_{A'}, \mathfrak{n}_{B'})
\end{array}$$

For $R^{\sqsubseteq}$, we obtain $\mathfrak{n} \rightarrow \mathfrak{n}_\top$ for all $\mathfrak{n} \in \mathcal{N}$ and

$$\begin{array}{rl}
\mathfrak{n}_{A''} \rightarrow A'' & \mathfrak{n}_\top \rightarrow \top \\
\mathfrak{n}_{A'} \rightarrow A' & \mathfrak{n}_{A'} \rightarrow \mathfrak{n}_B \\
\mathfrak{n}_A \rightarrow \mathfrak{n}_{A''} & \mathfrak{n}_{A'} \rightarrow \mathfrak{n}_A \\
\mathfrak{n}_A \rightarrow \mathfrak{n}_B & \mathfrak{n}_{A'} \rightarrow \mathfrak{n}_{A''} \\
\mathfrak{n}_{B'} \rightarrow \mathfrak{n}_A & \mathfrak{n}_{B'} \rightarrow \mathfrak{n}_{A''} \\
\mathfrak{n}_{B'} \rightarrow \exists s(\mathfrak{n}_A) & \mathfrak{n}_B \rightarrow \exists r(\mathfrak{n}_A) \\
\mathfrak{n}_A \rightarrow \exists r(\mathfrak{n}_A) & \mathfrak{n}_{A'} \rightarrow \exists r(\mathfrak{n}_A) \\
\mathfrak{n}_{B'} \rightarrow \exists s(\mathfrak{n}_\top) & \mathfrak{n}_B \rightarrow \exists r(\mathfrak{n}_\top) \\
\mathfrak{n}_A \rightarrow \exists r(\mathfrak{n}_\top) & \mathfrak{n}_{A'} \rightarrow \exists r(\mathfrak{n}_\top)
\end{array}$$

Additionally, $R^{\sqsubseteq}$ contains rules for conjunctions of all elements of $\mathtt{Post}_{Base}$ corresponding to $(\mathbf{GR3})$, which we do not give for space reasons.

By applying the rules $\mathfrak{n}_A \rightarrow \mathfrak{n}_{B'}, \mathfrak{n}_{B'} \rightarrow \exists s(\mathfrak{n}_A)$ contained in $R^{\sqsupseteq}$ n times, we obtain a term $\exists s(\exists s(...\exists s(A)))$ of depth n, which represents the corresponding subsumee of A of the same depth.

5.1 Grammar Properties

The following theorem states that the grammars derive only terms representing Σ-subsumees and Σ-subsumers of the corresponding atomic concept.

Theorem 2 *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox, Σ a signature and $A \in sig_C(\mathcal{T})$.*

1. *For each $t \in L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$, there is a concept C with $t_C = t$ and $sig(C) \subseteq \Sigma$ such that $\mathcal{T} \models C \sqsubseteq A$.*
2. *For each $t \in L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$, there is a concept C with $t_C = t$ and $sig(C) \subseteq \Sigma$ such that $\mathcal{T} \models A \sqsubseteq C$.*

Proof Sketch. The theorem is proved by an easy induction on the maximal nesting depth of functions in t using the rules given in Definition 2. $\square$

As discussed above, for the completeness of the grammar generating subsumees, we only guarantee to capture all associative variants of concepts not being obtained by adding arbitrary conjuncts to arbitrary sub-expressions.

Theorem 3 *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox, Σ a signature and $A \in sig_C(\mathcal{T})$.*

1. *For each C with $sig(C) \subseteq \Sigma$ such that $\mathcal{T} \models C \sqsubseteq A$ there is a concept C' such that C can be obtained from C' by adding arbitrary conjuncts to arbitrary sub-expressions and $t_{C'} \in L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$.*
2. *For each D with $sig(D) \subseteq \Sigma$ such that $\mathcal{T} \models A \sqsubseteq D$ holds: $t_D \in L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, A))$.*

Proof Sketch. The theorem is proved by induction on the role depth of C using the properties of the normalization, for instance, stated in Lemmas 2, in addition to Definition 2. $\square$

6 From Grammars to Uniform Interpolants

For the construction of a uniform interpolant, we make use of the results stated in Lemma 2, which, in combination with the introduced normalization imply that, knowing the subsumees and subsumers of atomic concepts in normalized terminologies is sufficient to derive all subsumptions between any complex concepts. In order to obtain a corresponding TBox from a pair of grammars, for all $\mathfrak{n}_B$ occurring on the right-hand sides of the transition rules must hold: $B \in \Sigma \cup \{\top\}$. If the latter is the case, we can apply the inverse substitution $\sigma^-(t)$ to obtain axioms defining subsumers and subsumees of atomic concepts. Otherwise, we first need to eliminate all non-terminals not from $\mathcal{N}^\Sigma = \{\mathfrak{n}_B \mid B \in \Sigma \cup \{\top\}\}$ within the right-hand sides of the corresponding rules. In principle, we can substitute any such non-terminal $\mathfrak{n} \notin \mathcal{N}^\Sigma$ by the right-hand sides of the corresponding rules for $\mathfrak{n}$ without any change to the generated language. However, in the general case, such a sequence of substitutions does not have to be finite. In the following, we investigate the bounds for the number of such substitution steps required to obtain a uniform interpolant.

For a concept C, let $d(C)$ denote the maximal role depth within C. For a TBox $\mathcal{T}$, $d(\mathcal{T}) = \mathtt{max}\{d(C) \mid C$ is a sub-expression of $\mathcal{T}\}$. The following lemma postulates a bound on the role depth of minimal uniform $\mathcal{EL}$ interpolants:

Lemma 3 *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox, Σ a signature. Let def$(\mathcal{T})$ be the number of definitions in $\mathcal{T}$. The following statements are equivalent:*

1. *There exists a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}$.*
2. *There exists a uniform $\mathcal{EL}$ Σ-interpolant $\mathcal{T}'$ of $\mathcal{T}$ and $d(\mathcal{T}') \leq 2^{4\cdot(|sig_C(\mathcal{T})|+def(\mathcal{T}))} + 1$.*

Proof Sketch. In a normalized TBox $\mathcal{T}$, the number of sub-expressions[2] is $|\text{sig}_C(\mathcal{T})| + \text{def}(\mathcal{T})$. Therefore, we can replace the last statement of Condition 2 by $d(\mathcal{T}') \leq 2^{2\cdot n} + 1$, where n is twice the number of sub-expressions within $\mathcal{T}$. Then, the lemma follows from Conditions (1) and (4) of Lemma 55 in [9]. □

We can eliminate all non-terminals not from $\mathcal{N}^\Sigma$ within the given role depth by replacing them in each rule by the corresponding right-hand sides, thereby obtaining a set of grammars that can be transformed into a uniform $\mathcal{EL}$ Σ-interpolant using the inverse substitution $\sigma^-(t)$.

Definition 3 *For a normalized $\mathcal{EL}$ TBox $\mathcal{T}$ and a signature Σ, let*

- $R_0^{\sqsupseteq} = R^{\sqsupseteq}$ *and* $R_0^{\sqsubseteq} = R^{\sqsubseteq}$.
- $R_{i+1}^{\bowtie} = \{\mathfrak{n} \to t(t_1', ..., t_n') \mid \mathfrak{n} \to t(\mathfrak{n}_1, ..., \mathfrak{n}_n) \in R_i^{\bowtie}, 1 \leq j \leq n, t_j' = \mathfrak{n}_j$ *if* $\mathfrak{n}_j \in \mathcal{N}^\Sigma$ *and* $t_j' \in \{t' \mid \mathfrak{n}_j \to t' \in R_0^{\bowtie}\}$ *for* $\mathfrak{n}_j \notin \mathcal{N}^\Sigma\}$ *with* $\bowtie \in \{\sqsupseteq, \sqsubseteq\}$.

For an $A \in sig_C(\mathcal{T})$, let $G_i^{\sqsupseteq} = (\mathfrak{n}_A, \mathcal{N}^\mathcal{T}, \mathcal{F}, R_i^{\sqsupseteq})$ and $G_i^{\sqsubseteq} = (\mathfrak{n}_A, \mathcal{N}^\mathcal{T}, \mathcal{F}, R_i^{\sqsubseteq})$. $\mathbb{G}_i^{\sqsupseteq}(\mathcal{T}, \Sigma)$ is then given by $\{G_i^{\sqsupseteq}(\mathcal{T}, \Sigma, A) \mid A \in sig_C(\mathcal{T})\}$ and $\mathbb{G}_i^{\sqsubseteq}(\mathcal{T}, \Sigma)$ by $\{G_i^{\sqsubseteq}(\mathcal{T}, \Sigma, A) \mid A \in sig_C(\mathcal{T})\}$.

Let $N = 2^{4\cdot(|\text{sig}_C(\mathcal{T})|+\text{def}(\mathcal{T}))} + 1$. Given a pair of grammar sets $\mathbb{G}_N^{\sqsupseteq}(\mathcal{T}, \Sigma), \mathbb{G}_N^{\sqsubseteq}(\mathcal{T}, \Sigma)$ for a TBox $\mathcal{T}$ and a signature Σ, we can compute a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}$ as follows.

[2] In a conjunction, only the concepts not being a conjunction itself are considered as proper sub-expressions. Therefore, a conjunction with n elements has n proper sub-expressions.

Definition 4 *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$ TBox, Σ a signature and $N = 2^{4\cdot(|sig_C(\mathcal{T})|+def(\mathcal{T}))} + 1$. Further, let $\mathbb{G}_1 = \mathbb{G}_N^{\sqsupseteq}(\mathcal{T}, \Sigma), \mathbb{G}_2 = \mathbb{G}_N^{\sqsubseteq}(\mathcal{T}, \Sigma)$ with $R_1 = R_N^{\sqsupseteq}$ and $R_2 = R_N^{\sqsubseteq}$. Then, $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) =$*

$$\{\sigma^-(t) \sqsubseteq A \mid A \in \Sigma, \mathfrak{n}_A \to t \in R_1, t \in T(\mathcal{F}, \mathcal{N}^\Sigma)\} \cup$$
$$\{A \sqsubseteq \sigma^-(t) \mid A \in \Sigma, \mathfrak{n}_A \to t \in R_2, t \in T(\mathcal{F}, \mathcal{N}^\Sigma)\} \cup$$
$$\{\sigma^-(t_1) \sqsubseteq \sigma^-(t_2) \mid \mathfrak{n} \notin \mathcal{N}_\Sigma, \mathfrak{n} \to t_1 \in R_1, \mathfrak{n} \to t_2 \in R_2,$$
$$t_1, t_2 \in T(\mathcal{F}, \mathcal{N}^\Sigma)\}.$$

Clearly, the construction terminates, if $\mathbb{G}_1$ and $\mathbb{G}_2$ are finite. The size of the resulting TBox $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma)$ is bounded polynomially by the size of $\mathbb{G}_1, \mathbb{G}_2$. Moreover, $\text{sig}(\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma)) \subseteq \Sigma$, since each $t, t_1, t_2 \in T(\mathcal{F}, \mathcal{N}^\Sigma)$, $\sigma^-(t) \subseteq \text{sig}(\mathcal{T})$ and $\mathcal{F} \cap (\text{sig}(\mathcal{T}) \setminus \Sigma) = \emptyset$. We obtain the following result concerning the size of uniform $\mathcal{EL}$ Σ-interpolants of $\mathcal{T}$.

Theorem 4 *Let $\mathcal{T}$ be an $\mathcal{EL}$ TBox and Σ a signature. The following statements are equivalent:*

1. *There exists a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}$.*
2. $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \equiv_\Sigma^c \mathcal{T}$
3. *There exists a uniform $\mathcal{EL}$ Σ-interpolant $\mathcal{T}'$ with $|\mathcal{T}'| \in O(2^{2^{2^{|\mathcal{T}|}}})$.*

Proof. The non-trivial parts of the proof are implications $1 \Rightarrow 2$ and $2 \Rightarrow 3$.

$1 \Rightarrow 2$**:** By Definition 1, the statement $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \equiv_\Sigma^c \mathcal{T}$ consists of two directions: (1) for all $\mathcal{EL}$ concepts C, D with $\text{sig}(C) \cup \text{sig}(D) \subseteq \Sigma$ holds $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models C \sqsubseteq D \Rightarrow \mathcal{T} \models C \sqsubseteq D$ and (2) for all $\mathcal{EL}$ concepts C, D with $\text{sig}(C) \cup \text{sig}(D) \subseteq \Sigma$ holds $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models C \sqsubseteq D \Leftarrow \mathcal{T} \models C \sqsubseteq D$.

(1) The first direction follows from Theorem 2 and Definition 4, which does not introduce any consequences not being consequences of $\mathcal{T}$.

(2) For the second direction, assume that there exists a uniform $\mathcal{EL}$ Σ-interpolant of $\mathcal{T}$. Then, by Lemma 3, there exists a uniform $\mathcal{EL}$ Σ-interpolant $\mathcal{T}'$ of $\mathcal{T}$ with $d(\mathcal{T}') \leq N$. It is sufficient to show that for each $C \sqsubseteq D \in \mathcal{T}'$ holds $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models C \sqsubseteq D$. Assume that $C \sqsubseteq D \in \mathcal{T}'$. Then, $\mathcal{T} \models C \sqsubseteq D$ and we prove by induction on maximal role depth of C, D that also $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models C \sqsubseteq D$. W.l.o.g., let $D = \bigsqcap_{1\leq i\leq l} D_i$ and

$$C = \bigsqcap_{1\leq j\leq n} A_j \sqcap \bigsqcap_{1\leq k\leq m} \exists r_k.E_k$$

with $A_j \in \Sigma \cap \text{sig}_C(\mathcal{T})$ for $1 \leq j \leq n$, $r_k \in \Sigma \cap \text{sig}_R(\mathcal{T})$ for $1 \leq k \leq m$ and E_k with $1 \leq k \leq m$ a set of $\mathcal{EL}$ concepts such that $\text{sig}(E_k) \subseteq \Sigma$. Clearly, $\mathcal{T} \models C \sqsubseteq D$, iff $\mathcal{T} \models C \sqsubseteq D_i$ for all i with $1 \leq i \leq l$.

- If $D_i = A \in \Sigma$, then, it follows from Theorem 3 that there is a concept C' such that C can be obtained from C' by adding arbitrary conjuncts to arbitrary sub-expressions with $t_{C'} \in L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$. Since $d(C) \leq N$ and C has been obtained from C' by weakening, also $d(C') \leq N$. Therefore, $t_{C'} \in L(G_N^{\sqsupseteq}(\mathcal{T}, \Sigma, A))$, and $\mathtt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models C \sqsubseteq D_i$.
- If $D_i = \exists r.D'$ for some r, D', then, by Lemma 2, one of the following is true:

(A3) There are r_k, E_k in C such that $r_k = r$ and $\mathcal{T} \models E_k \sqsubseteq D'$. Since $d(E_k) < N$ and $d(D') < N$, by induction hypothesis holds $\texttt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models E_k \sqsubseteq D'$. It follows that $\texttt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models \exists r_k.E_k \sqsubseteq D_i$ and $\texttt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models C \sqsubseteq D_i$.

(A4) There is $B \in \mathrm{sig}_C(\mathcal{T})$ of $\mathcal{T}$ such that $\mathcal{T} \models B \sqsubseteq \exists r.D'$ and $\mathcal{T} \models C \sqsubseteq B$. Then,

- it follows from Theorem 3 that there is a concept C' such that C can be obtained from C' by adding arbitrary conjuncts to arbitrary sub-expressions with $t_{C'} \in L(G^{\sqsupseteq}(\mathcal{T}, \Sigma, B))$. Since $d(C) \leq N$ and C has been obtained from C' by weakening, also $d(C') \leq N$. Therefore, $t_{C'} \in L(G^{\sqsupseteq}_N(\mathcal{T}, \Sigma, B))$
- it follows from Theorem 3 that $t_{\exists r.D'} \in L(G^{\sqsubseteq}(\mathcal{T}, \Sigma, B))$. Since $d(\exists r.D') \leq N$, it follows that $t_{\exists r.D'} \in L(G^{\sqsubseteq}_N(\mathcal{T}, \Sigma, B))$.

Therefore, by Definition 4, $\texttt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models C' \sqsubseteq \exists r.D'$, and $\texttt{UI}(\mathbb{G}_1, \mathbb{G}_2, \Sigma) \models C \sqsubseteq D_i$.

$2 \Rightarrow 3$**:** Observe that $\mathbb{G}_1, \mathbb{G}_2$ have $|\mathrm{sig}_C(\mathcal{T})|$ non-terminals and at most $2^{2\cdot n} + |\mathrm{sig}_C(\mathcal{T})|$ outgoing transitions for each non-terminal, n the maximal arity of $\sqcap$, each of which has at most n occurring non-terminals. Let $\texttt{leaves}_i$ be the maximal number of non-terminals $\mathfrak{n} \notin \mathcal{N}^\Sigma$ occurring in a transition after step i and $\texttt{tran}_i$ the maximal number of outgoing transitions for a non-terminal after step i. Then, $\texttt{tran}_0 = 2^{2\cdot n} + |\mathrm{sig}_C(\mathcal{T})|$ and $\texttt{leaves}_0 = n$. Further, $\texttt{leaves}_{i+1} = n \cdot \texttt{leaves}_i$, i.e., $\texttt{leaves}_i = n^{i+1}$. For each $\mathfrak{n} \notin \mathcal{N}^\Sigma$, there are at most $2^{2\cdot n} + |\mathrm{sig}_C(\mathcal{T})|$ possible replacing transitions, therefore, for each $t \in R_i$, there are $(2^{2\cdot n} + |\mathrm{sig}_C(\mathcal{T})|)^{\texttt{leaves}_{i+1}}$ possibilities to replace all non-terminals $\mathfrak{n} \notin \mathcal{N}^\Sigma$ by the corresponding transitions from R_0. We obtain $\texttt{tran}_{i+1} = \texttt{tran}_i \cdot (2^{2\cdot n} + |\mathrm{sig}_C(\mathcal{T})|)^{\texttt{leaves}_{i+1}}$, i.e., $\texttt{tran}_i \leq (2^{2\cdot n} + |\mathrm{sig}_C(\mathcal{T})|)^{i\cdot n^{i+2}}$. For $i = N$, we obtain $\texttt{leaves}_i = n^N \in O(2^{2^{|\mathcal{T}|}})$ and $\texttt{tran}_i \leq (2^{2\cdot n} + |\mathrm{sig}_C(\mathcal{T})|)^{(N)\cdot n^{N+2}} \in O(2^{2^{2^{|\mathcal{T}|}}})$.

These complexity results correspond to the size and number of axioms in Example 2. □

7 Summary and Future Work

In this paper, we provide an approach to computing uniform interpolants of general $\mathcal{EL}$ terminologies based on proof theory and regular tree languages. Moreover, we show 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.

Due to the triple exponential blowup, algorithms for testing the appropriate size of uniform interpolants in addition to their existence would be of importance for applications in practice. While, in principle, expressing uniform interpolants in $\mathcal{EL}$ extended with fixpoint constructs [13] allows us to avoid both problems, the non-existence and the triple exponential blowup, for practical scenarios, reducing the forgotten signature in a reasonable way would be an interesting alternative, for instance, for applications as visualization of dependencies or ontology reuse.

Moreover, given the considerable effect of structure sharing elimination on the size of a TBox, it would be interesting to investigate, to what extent the structure sharing within existing large ontologies can be intensified in order to make reasoning more efficient.

ACKNOWLEDGEMENTS

This work was supported by the project ExpresST funded by the German Research Foundation (DFG).

REFERENCES

[1] F. Baader, S. Brandt, and C. Lutz, 'Pushing the $\mathcal{EL}$ envelope', in *Proc. of the 19th Int. Joint Conf. on Artificial Intelligence (IJCAI-05)*, pp. 364–369, (2005).

[2] *The Description Logic Handbook: Theory, Implementation, and Applications*, eds., Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter Patel-Schneider, Cambridge University Press, second edn., 2007.

[3] Sebastian Brandt, 'Polynomial time reasoning in a description logic with existential restrictions, gci axioms, and - what else?', in *Proc. of the 16th European Conf. on Artificial Intelligence (ECAI-04)*, pp. 298–302, (2004).

[4] Simona Colucci, Tommaso Di Noia, Eugenio Di Sciascio, Francesco M. Donini, and Azzurra Ragone, 'A unified framework for non-standard reasoning services in description logics', in *Proc. of the 19th European Conf. on Artificial Intelligence (ECAI-10)*, pp. 479–484, (2010).

[5] Hubert Comon, Florent Jacquemard, Max Dauchet, Remi Gilleron, Denis Lugiez, Christof Loding, Sophie Tison, and Marc Tommasi, *Tree Automata Techniques and Applications*, 2008.

[6] Yevgeny Kazakov, 'Consequence-driven reasoning for $\mathcal{H}$orn $\mathcal{SHIQ}$ ontologies', in *Proc. of the 21st Int. Joint Conf. on Artificial Intelligence (IJCAI-09)*, pp. 2040–2045, (2009).

[7] Boris Konev, Dirk Walther, and Frank Wolter, '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).

[8] Roman Kontchakov, Frank Wolter, and Michael Zakharyaschev, 'Logic-based ontology comparison and module extraction, with an application to dl-lite', *Artif. Intell.*, **174**, 1093–1141, (October 2010).

[9] Carsten Lutz, Inanc Seylan, and Frank Wolter, '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-2012)*, (2012).

[10] Carsten Lutz and Frank Wolter, 'Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$', *Journal of Symbolic Computation*, **45**(2), 194–228, (2010).

[11] Carsten Lutz and Frank Wolter, '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).

[12] *OWL 2 Web Ontology Language: Profiles*, eds., Boris Motik, Bernardo Cuenca Grau, Ian Horrocks, Zhe Wu, Achille Fokoue, and Carsten Lutz, W3C Recommendation, 27 October 2009. Available at http://www.w3.org/TR/owl2-profiles/.

[13] Nadeschda Nikitina, 'Forgetting in general el terminologies', in *Proc. of the 24th Int. Workshop on Description Logics (DL2011)*, (2011).

[14] Nadeschda Nikitina and Sebastian Rudolph, 'ExpExpExplosion: Uniform interpolation in general el terminologies', Technical report, AIFB, KIT, Karlsruhe, (May 2012).

[15] Nadeschda Nikitina, Sebastian Rudolph, and Birte Glimm, 'Reasoning-supported interactive revision of knowledge bases', in *Proc. of the 22nd Int. Joint Conf. on Artificial Intelligence (IJCAI-11)*, pp. 1027–1032, (2011).

[16] W3C OWL Working Group, *OWL 2 Web Ontology Language: Document Overview*, W3C Recommendation, 27 October 2009. Available at http://www.w3.org/TR/owl2-overview/.

[17] Sebastian Rudolph, 'Foundations of description logics', in *Reasoning Web*, pp. 76–136, (2011).