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

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 [15], description logics [2, 16] 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, 14].

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, 14].

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 non-terminal 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 4 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 5. In the same section, we also show the upper bound on the size of uniform interpolants. We summarize the contributions in Section 6 and discuss some ideas for future work. Detailed proofs are available in the extended version of this paper.

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

$$C := A \mid T \mid |C \cap C| \exists r.C$$

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 $C \subseteq D$ and $D \subseteq C$. 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 $\alpha$, denoted by sig($C$) or 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 $\text{sig}_C(C)$ and $\text{sig}_R(C)$, respectively. The signature of a TBox $\mathcal{T}$, in symbols $\text{sig}(\mathcal{T})$ (correspondingly, $\text{sig}_C(\mathcal{T})$ and $\text{sig}_R(\mathcal{T})$), 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 to each concept $A \in NC$ a subset $A^\mathcal{I}$ of $\Delta^\mathcal{I}$ and each role $r \in NR$ a subset $r^\mathcal{I}$ of $\Delta^\mathcal{I} \times \Delta^\mathcal{I}$. The interpretation of $\mathcal{I}$ is fixed to $\Delta^\mathcal{I}$. An interpretation of an arbitrary $\mathcal{EL}$ concept is defined inductively, i.e., $(C \cap D)^\mathcal{I} = C^\mathcal{I} \cap D^\mathcal{I}$ and $(\exists r.C)^\mathcal{I} = \{ x \mid (x, y) \in \Delta^\mathcal{I} \times \Delta^\mathcal{I}, y \in C^\mathcal{I} \}$. An interpretation $\mathcal{I}$ satisfies an axiom $C \subseteq D$ 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 $\alpha$ (in symbols, $\mathcal{T} \models \alpha$), if $\alpha$ is satisfied by all models of $\mathcal{T}$.

Tree Languages and Regular Tree Grammars

A ranked alphabet is a pair $(\mathcal{F}, \text{Arity})$ where $\mathcal{F}$ is a finite set and Arity is a mapping from $\mathcal{F}$ into $\mathbb{N}$. $\mathcal{T}(\mathcal{F})$ denotes the set of ground terms over the alphabet $\mathcal{F}$. Let $X_n$ be a set of $n$ variables. A term $C \in \mathcal{T}(\mathcal{F}, X_n)$ containing each variable from $X_n$ at most once is called a context. We denote by $\mathcal{C}(\mathcal{F})$ the set of contexts containing a single variable. A regular tree grammar $G = (S, \mathcal{F}, \mathcal{R}, R)$ is composed of a start symbol $S$, a set $\mathcal{F}$ of non-terminal symbols (non-terminal symbols have arity 0) with $S \in \mathcal{F}$, 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 $X \rightarrow \beta$ where $\beta$ is a tree of $\mathcal{T}(\mathcal{F} \cup \mathcal{N})$ and $X \in \mathcal{N}$. Given a regular tree grammar $G = (S, \mathcal{F}, \mathcal{R}, R)$, the derivation relation $\rightarrow_G$ associated to $G$ is a relation on pairs of terms of $\mathcal{T}(\mathcal{F} \cup \mathcal{N})$ such that $s \rightarrow_G t$ if and only if there is a rule $X \rightarrow \alpha \in R$ 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 $\mathcal{T}(\mathcal{F})$ which can be reached by successive derivations starting from the start symbol, i.e., $L(G) = \{ s \in \mathcal{T} \mid s \rightarrow^+ s \} \cup \rightarrow^+$ the transitive closure of $\rightarrow$. We write $\rightarrow$ instead of $\rightarrow_G$ when the grammar $G$ is clear from the context. For further details, we refer the reader, for instance, to [3].

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 $\Sigma$ if they have the same $\Sigma$-consequences, i.e., consequences whose signature is a subset of $\Sigma$. Depending on the particular application requirements, the expressivity of these $\Sigma$ 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., [3] or [7]. As follows:

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

Given a signature $\Sigma$ and a TBox $\mathcal{T}$, the aim of uniform interpolation is to determine a TBox $\mathcal{T}'$ with $\text{sig}(\mathcal{T}') \subseteq \Sigma$ such that $\mathcal{T} \equiv_\Sigma \mathcal{T}'$. $\mathcal{T}'$ is also called a uniform $\mathcal{EL}$ $\Sigma$-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 $\mathcal{EL}$ $\Sigma$-interpolant might not exist for a particular TBox $\mathcal{T}$ and a particular $\Sigma$.

**Example 1** Consider uniform interpolants of the TBox $\mathcal{T} = \{ A' \sqsubseteq A, A, A' \sqsubseteq A', A \sqsubseteq 3r.A, 3s.A \sqsubseteq A \}$, w.r.t. $\Sigma = \{ s, r, A', A'' \}$. We obtain an infinite chain of consequences $A' \sqsubseteq 3r.A, 3r.3r.A', 3r.3r.3r.A''$, and consequently $\exists r.3s.3s.3s.A'' \sqsubseteq A'$ 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 $\mathcal{AC}_I$ [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 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:

\[
\begin{align*}
A_1 & \subseteq X_0 \sqcap \ldots \sqcap X_{n-1} \\
A_2 & \subseteq X_0 \sqcap \ldots \sqcap X_{n-1} \\
\forall_{\sigma \in \{r,s\}} \exists \sigma. (X_i \sqcap X_0 \sqcap \ldots \sqcap X_{i-1}) & \subseteq X_i \quad i < n \\
\forall_{\sigma \in \{r,s\}} \exists \sigma. (X_i \sqcap X_0 \sqcap \ldots \sqcap X_{i-1}) & \subseteq X_i \quad i < n \\
\forall_{\sigma \in \{r,s\}} \exists \sigma. (X_i \sqcap X_0 \sqcap \ldots \sqcap X_{i-1}) & \subseteq X_i \quad j < i < n \\
X_0 \sqcap \ldots \sqcap X_{n-1} & \subseteq B 
\end{align*}
\]

If we now consider sets $C_i$ of concept descriptions inductively defined by $C_0 = \{ A_1, A_2 \}, C_{i+1} = \{ 3r.C_i \sqcap 3s.C_2 \mid C_i, C_2 \in C_i \}$, then we find that $|C_{n+1}| = |C|^2$ and consequently $|C_n| = 2^{|C|^2}$. Thus, the set $C_{2n+1}$ contains triply exponentially many different concepts, each of which is doubly exponential in the size of $\mathcal{T}_n$ (intu- itively, 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). It is straightforward to check that for each concept $C \in C_{2n+1}$ 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 Appendix B).

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 [15, 21], 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 $C_1 \sqcap \ldots \sqcap C_n \sqcap \exists r.C$ by fresh concept symbols until each axiom in the TBox $\mathcal{T}$ is one of $\{ A \sqsubseteq B, A \equiv B, A \equiv_1 \ldots \equiv_n B, A \equiv 3r.B \}$, where $A, B, r, s \in \text{sig}(\mathcal{T}) \cup \{ \top \}$ and $r \in \text{sig}(\mathcal{T})$. For this purpose, we introduce a minimal required set of fresh concept symbols
is then given by

\[ \exists n \text{ for each } n \in \mathbb{N} \]

Let \( R^E \) be given by:

\begin{align*}
(GR1) & \ni_T \rightarrow \top \\
& \text{for all } B \in \Sigma_C
\end{align*}

Proof Sketch. For Example 3

Example 3 For \( T \) and \( \Sigma \) from Example 1 we obtain a normalized TBox \( T' = \{ A' \subseteq A, A \subseteq A', A \subseteq B, B \equiv \exists n A, A' \equiv \exists s A, B' \subseteq A \} \), which yields the following set of transitions for \( R^E \):

\begin{align*}
n_T & \rightarrow \n(n_T, n_T) \\
n_T & \rightarrow \exists n(n_T) \\
n_T & \rightarrow \exists s(n_T) \\
n_A & \rightarrow A'' \\
n_A & \rightarrow \exists n_A \\
n_A & \rightarrow n_A \\
n_B & \rightarrow \exists n_B \\
n_B & \rightarrow n_B \\
n_B & \rightarrow \exists s(n_B) \\
n_B & \rightarrow \exists s(n_A)
\end{align*}

For \( R^E \), we obtain

\begin{align*}
n_T & \rightarrow \top \\
n_A & \rightarrow n_T \\
n_B & \rightarrow n_T \\
n_B & \rightarrow n_A \\
n_B & \rightarrow \exists s(n_A) \\
n_B & \rightarrow \exists s(n_T)
\end{align*}

By applying the rules \( n_A \rightarrow n_B, n_B \rightarrow \exists s(n_A) \) contained in \( R^E \) \( n \) times, we obtain a term \( \exists s(\exists s(...(\exists s(n_A))) \) of depth \( n \), which represents the corresponding subsume of \( A \) of the same depth.

We enrich the rules as shown by the following definition in order to extend the generated languages by associative variants of concept expressions. For this purpose, we consider subsumees and subsumers of each atomic concept having the form of simple conjunctions, i.e., conjunctions of simple concepts. While, in the case of subsumees (Pre(A)) it is sufficient to consider atomic concepts only, in the case subsumers (Post(A)), we additionally have to take into account existential restrictions with atomic concepts to account for the corresponding associative variants.

**Grammar Properties**

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

**Theorem 2** Let \( T \) be a normalized EL TBox, \( \Sigma \) a signature and \( A \in \Sigma_C(T) \).

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

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

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

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 subexpressions (ANDL-weakening, Figure 1). The reason for this limitation is that, in general, adding arbitrary conjuncts to arbitrary subexpressions allows us to obtain subsumees being conjunctions of unbounded size, which would cause the corresponding language to contain terms with \( \gamma \)-functions of unbounded arity and make the definition of the grammar unnecessary complex. We show in the next section that the subset of subsumees covered by the grammar is sufficient to preserve all \( \Sigma \) subsumees.

**Theorem 3** Let \( T \) be a normalized EL TBox, \( \Sigma \) a signature and \( A \in \Sigma_C(T) \).

1. For each \( C \) with \( \text{sig}(C) \subseteq \Sigma \) such that \( T \models C \subseteq A \) there is a concept \( C' \) such that \( C \) can be obtained from \( C' \) by adding arbitrary conjuncts to arbitrary subexpressions and \( t_{C'} \in L(G^E(T), A) \).

2. For each \( D \) with \( \text{sig}(D) \subseteq \Sigma \) such that \( T \models A \subseteq D \) holds: \( t_D \in L(G^E(T), \Sigma, A) \).

Proof Sketch. The theorem is proved by induction on the role depth of \( C \) using Lemmas 2 and 4 in addition to Definitions 2 and 4.
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 subsumers and subsumers of atomic concepts in normalized terminologies is sufficient to derive all subsumptions between any complex concepts. This justifies the computation of the uniform interpolant based on the grammars introduced in the last section. In order to obtain a corresponding TBox from a pair of grammars, for all $m_B$ occurring on the right-hand sides of the transition rules must hold: $B \in \Sigma \cup \{T\}$. If the latter is the case, we can apply the inverse substitution $\sigma^{-1}(t)$ to obtain axioms defining subsumers and subsumers of atomic concepts.

Otherwise, we first need to eliminate all non-terminals not from $\Lambda^2 = \{B \mid B \in \Sigma \cup \{T\}\}$ within the right-hand sides of the corresponding rules. In principle, we can substitute any such non-terminal $N \not\in \Lambda^2$ by the right-hand sides of the corresponding rules for $N$ 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 an uniform interpolant.

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

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

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

**Proof Sketch.** In a normalized TBox $T$, the number of subconcepts is $|\text{sig}(T)| + \text{def}(T)$. Therefore, we can replace the last statement from Condition 2 by $d(T') \leq 2^{2^n} + 1$, where $n$ is twice the number of subconcepts within $T$. Then, the lemma follows from Conditions (1) and (4) of Lemma 55 in [3].

We can eliminate all non-terminals not from $\Lambda^2$ 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 $\mathcal{EL}$ $\Sigma$-interpolant using the inverse substitution $\sigma^{-1}(t)$.

**Definition 5** For a normalized $\mathcal{EL}$ TBox $T$ and a signature $\Sigma$, let

- $R_0^2 = R_0$ and $R_0^C = R_0$.
- $R_{2i}^0 = \{N \rightarrow t(1,\ldots,t_n) \mid N \rightarrow t(N_1,\ldots,N_n) \in R_{2i-1}^0, 1 \leq j \leq n, t'_j \in N_j \subseteq \Lambda^2 \}$ for $i > 0$.
- $R_{2i-1} = \{N \rightarrow t(1,\ldots,t_n) \mid N \rightarrow t(N_1,\ldots,N_n) \in R_{2i-2}^0, 1 \leq j \leq n, t'_j \in \{t_i \mid N_j = \Lambda^2 \}$ for $i > 0$.

For an $A \in \text{sig}(T)$, let $G_A^2 = (\Sigma,\Lambda^2,\mathcal{F},R_0^2)$ and $G_A^C = (\Sigma,\Lambda^C,\mathcal{F},R_0^C)$, where $G_A^C(T,\Sigma,\Lambda) \subseteq G_A^C(T,\Sigma,\Lambda)$. Then, $G_A^2(T,\Sigma,\Lambda) \subseteq G_A^2(T,\Sigma,\Lambda)$ is a uniform $\mathcal{EL}$ $\Sigma$-interpolant of $T$.

**Definition 6** Let $T$ be a normalized $\mathcal{EL}$ TBox, $\Sigma$ a signature and $G_1 = G_{1}^{\mathcal{F}}(\Lambda \cup \{t(1,\ldots,t_n) \mid N \rightarrow t(N_1,\ldots,N_n) \in R_{2i-1}^0, 1 \leq j \leq n, t'_j \in \{t_i \mid N_j = \Lambda^2 \}$ for $i > 0$). Then, $G_2$ is a uniform $\mathcal{EL}$ $\Sigma$-interpolant of $T$.

**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 $U \Sigma_1(G_1, G_2, \Sigma) \equiv_\Sigma T$ consists of two directions: (1) for all $\mathcal{EL}$ concepts $C, D$ with $\text{sig}(C) \cup \text{sig}(D) \subseteq \Sigma$ holds $U \Sigma_1(G_1, G_2, \Sigma) \models C \subseteq D \Rightarrow T \models C \subseteq D$ and (2) for all $\mathcal{EL}$ concepts $C, D$ with $\text{sig}(C) \cup \text{sig}(D) \subseteq \Sigma$ holds $U \Sigma_1(G_1, G_2, \Sigma) \models C \subseteq D \Leftrightarrow T \models C \subseteq D$.

(1) The first direction follows from Theorem[2] and Definition[6] which does not introduce any consequencies not being consequencies of $T$.

(2) The second direction, assume that there exists a uniform $\mathcal{EL}$ $\Sigma$-interpolant of $T$. By Lemma[2] there exists a uniform $\mathcal{EL}$ $\Sigma$-interpolant $T'$ of $T$ with $d(T') \leq 2^{4^{|\text{sig}(T)| + \text{def}(T)|}} + 1$. It is sufficient to show that for each $C \subseteq D \models T'$ holds $U \Sigma_1(G_1, G_2, \Sigma) \models C \subseteq D$. Assume that $C \subseteq D \models T'$. Then, $T \models C \subseteq D$ and prove by induction on maximal role depth of $C, D$ that also $U \Sigma_1(G_1, G_2, \Sigma) \models C \subseteq D$. W.l.o.g., let $D = \bigcap_{1 \leq j \leq m} D_j$.

$$C = \bigcap_{1 \leq j \leq l} \bigcap_{1 \leq k \leq m} A_j \cap \exists \mathbb{E}_k$$

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

• 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 subexpressions with $t_{C'} \in L(G^3(T, \Sigma, A))$. Since $d(C') \leq 2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|) + 1$ and $C$ has been obtained from $C'$ by weakening, also $d(C') \leq 2^{4(|\text{sig}_{C'}(T)| + |\text{def}(T)|)} + 1$. Therefore, $t_{C'} \in L(G^3_{2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|)+1}(T, \Sigma, A))$, and $\mathcal{U}(G_1, G_2, \Sigma) \models C \subseteq D_i$.

- 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 $T \models E_k \subseteq D'$. Since $d(E_k) < 2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|) + 1$ and $d(D') < 2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|) + 1$, by induction hypothesis holds $\mathcal{U}(G_1, G_2, \Sigma) \models E_k \subseteq D'$. It follows that $\mathcal{U}(G_1, G_2, \Sigma) \models \exists r_k, E_k \subseteq D_i$ and $\mathcal{U}(G_1, G_2, \Sigma) \models C \subseteq D_i$.

(A4) There is $B \in \text{sig}_{C'}(T)$ of $T$ such that $T \models B \equiv \exists \exists. D'$, and $T \models C \subseteq B$. 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 subexpressions with $t_{C'} \in L(G^3(T, \Sigma, B))$. Since $d(C) \leq 2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|) + 1$ and $C$ has been obtained from $C'$ by weakening, also $d(C) \leq 2^{4(|\text{sig}_{C'}(T)| + |\text{def}(T)|)} + 1$. Therefore, $t_{C'} \in L(G^3_{2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|)+1}(T, \Sigma, B))$.

- it follows from Theorem 3 that $t_{\exists \exists, D'} \in L(G^3_{2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|)+1}(T, \Sigma, B))$. Since $d(\exists \exists, D') \leq 2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|) + 1$, it follows that $t_{\exists \exists, D'} \in L(G^3_{2^4(|\text{sig}_{C'}(T)| + |\text{def}(T)|)+1}(T, \Sigma, B))$.

Therefore, by Definition 6 $\mathcal{U}(G_1, G_2, \Sigma) \models C' \equiv \exists \exists. D'$, and $\mathcal{U}(G_1, G_2, \Sigma) \models C \subseteq D_i$.

2 $\Rightarrow$ 3: Observe that $G_1, G_2$ have $|\text{sig}_{C'}(T)|$ non-terminals and at most $2^2n + |\text{sig}_{C'}(T)|$ outgoing transitions for each non-terminal, $n$ the maximal arity of $\exists \exists$, each of which has at most $n$ occurring non-terminals. Let $\text{leaves}$, be the maximal number of non-terminals $N \notin \Sigma^2$ occurring in a transition after step $i$ and $\text{tran}_i$ the maximal number of outgoing transitions for a non-terminal after step $i$. Then, $\text{tran}_0 = 2^2n + |\text{sig}_{C'}(T)|$ and $\text{leaves}_0 = n$. Further, $\text{leaves}_{i+1} = n \cdot \text{leaves}_i$, i.e., $\text{leaves}_i = n^{i+1}$. For each $N \notin \Sigma^2$ there are at most $2^2n + |\text{sig}_{C'}(T)|$ possible replacing transitions, therefore, for each $t \in R_0$, there are $2^2n + |\text{sig}_{C'}(T)|$ $\text{leaves}_{i+1}$-possibilities to replace all non-terminals $N \notin \Sigma^2$ by the corresponding transitions from $R_0$. We obtain $\text{tran}_{i+1} = \text{tran}_i + (2^2n + |\text{sig}_{C'}(T)|)^{\text{leaves}_{i+1}}$, i.e., $\text{tran}_i \leq (2^2n + |\text{sig}_{C'}(T)|)^{i+1} + 1$, and we obtain $\text{leaves}_i = n^{2^2n + |\text{sig}_{C'}(T)|} \in \mathcal{O}(2^4|\text{sig}_{C'}(T)|)$.

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 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.

REFERENCES


A Proof Theory

The structure of the grammars has been derived based on Proof Theory. The used Gentzen-style proof system shown below has been derived similarly to the proof system for Horn-SKEL Q terminologies presented in [5]. In contrast to the proof system by Kazakov, which is complete for classification only and based on a normalization involving inverse rules (e.g., encoding all \( \exists x A \subseteq B \) as \( A \subseteq \forall x \neg B \)), the rules presented below fit our normal form and are complete for arbitrary \( \mathcal{EL} \) GCIs.

\[
\begin{align*}
C \subseteq C \quad & \text{(AX)} \\
D \subseteq E & \text{ (ANDL)} \\
C \subseteq D & \text{ (ANDR)} \\
C \subseteq E \quad & \text{(EX)} \\
C \subseteq D & \text{ (CUT)} \\
\end{align*}
\]

Figure 2. Gentzen-style proof system for general \( \mathcal{EL} \) terminologies.

Lemma 4 (Soundness and Completeness) Let \( \mathcal{T} \) be an arbitrary \( \mathcal{EL} \) TBox, \( C, D \in \mathcal{EL} \) concepts. Then \( \mathcal{T} \models C \subseteq D \) iff \( \mathcal{T} \models C \sqsubseteq D \).

Proof. While the soundness of the proof system (if-direction) can be easily checked for each rule, the proof of completeness is more sophisticated. In order to show the only-if-direction of the lemma, we will show that the following claim holds for \( \mathcal{T} \): For all \( \delta_E \in \Delta^\mathcal{T} \) and \( \mathcal{EL} \) concepts \( F \) holds \( \delta_E \in F^\mathcal{T} \) iff \( \mathcal{T} \models E \sqsubseteq F \).

We will show that the following claim holds for \( \mathcal{T} \): For all \( \delta_E \in \Delta^\mathcal{T} \) and \( \mathcal{EL} \) concepts \( F \) holds \( \delta_E \in F^\mathcal{T} \) iff \( \mathcal{T} \models E \sqsubseteq F \).

This claim can be exploited in two ways: First, we use it to show that \( \mathcal{T} \) is indeed a model of \( \mathcal{T} \). Let \( C \subseteq D \in \mathcal{T} \) and consider an arbitrary \( \delta_C \in \Delta^\mathcal{T} \) with \( \delta_C \in C^\mathcal{T} \). Via \((*)\) we obtain \( \mathcal{T} \models \top \subseteq C \) on the other hand, \( \mathcal{T} \models C \subseteq D \) due to \( C \subseteq D \in \mathcal{T} \). Thus we can derive \( \mathcal{T} \models \top \subseteq D \) via \( \text{(CUT)} \) and consequently, applying \((*)\) again, we obtain \( \delta_C \in D^\mathcal{T} \). Thereby modellhood of \( \mathcal{T} \) wrt. \( \mathcal{T} \) has been proven.

Second, we use \((*)\) to show that \( \mathcal{T} \) is a counter-model for all GCIs not derivable from \( \mathcal{T} \) as follows: Assume \( \mathcal{T} \models C \subseteq D \) but \( \mathcal{T} \not\models C \subseteq D \). Then \( \Delta^\mathcal{T} \) contains the element \( \delta_C \). From \( \mathcal{T} \models C \subseteq C \) and \((*)\) we derive \( \delta_C \in \Delta^\mathcal{T} \), from \( \mathcal{T} \not\models C \subseteq D \) and \((*)\) we obtain \( \delta_C \in D^\mathcal{T} \). Hence we get \( C^\mathcal{T} \subseteq D^\mathcal{T} \) and therefore \( \mathcal{T} \not\models C \subseteq D \), a contradiction.

It remains to prove \((*)\). This is done by an induction on the maximal nesting depth of the operators \( \exists \) and \( \forall \). There are two base cases:

- for \( F = \top \), the claim trivially follows from \( \text{(AXTOP)} \).
- for \( F \in \text{sig}_C(T) \), it is a direct consequence of the definition.

we now consider the cases where \( F \) is a complex concept expression

- for \( F = C_1 \cap \ldots \cap C_n \), we note that \( \delta_E \in F^\mathcal{T} \) exactly if \( \delta_E \in C_i^\mathcal{T} \) for all \( i \in \{1, \ldots, n\} \). By induction hypothesis, this means \( \mathcal{T} \models E \subseteq C_i \) for all \( i \in \{1, \ldots, n\} \). Finally, observe that \( \{E \subseteq \bigcup_{i \in \{1, \ldots, n\}} C_i \} \) can be mutually derived from each other: (for "\( \models \)" this is a straightforward consequence of (ANDR), for "\( \not\models \)" note that we can derive \( \emptyset \not\models C_i \subseteq C_i \) and \( \emptyset \not\models \bigcap_{i \in \{1, \ldots, n\}} C_i \subseteq C_i \) whence together with \( E \subseteq C_1 \cap \ldots \cap C_n \) follows \( E \subseteq C_i \) by \( \text{(CUT)} \).

- for \( F = \exists r.G \), we prove the two directions separately. First assuming \( \delta_E \in F^\mathcal{T} \) we must find \( (\delta_E, \delta_H) \in r^\mathcal{T} \) for some \( H \) with \( \delta_H \in G^\mathcal{T} \). This implies both \( \mathcal{T} \models E \subseteq \exists r.H \) (by definition) and \( \mathcal{T} \models \exists r.G \) (via the induction hypothesis). From the latter, we can deduce \( \mathcal{T} \models \exists r.H \subseteq \exists r.G \) by (Ex) and consequently \( \mathcal{T} \models E \subseteq \exists r.G \). For the other direction, note that by definition, \( \mathcal{T} \models E \subseteq \exists r.G \) implies \( (\delta_E, \delta_G) \in r^\mathcal{T} \). On the other hand, we get \( \mathcal{T} \models G \subseteq \top \) by (AX) and therefore \( \delta_G \in G^\mathcal{T} \) by the induction hypothesis which yields us \( \delta_E \in F^\mathcal{T} \).

B Proof of Lower Bound

Theorem 5 There exists a sequence of \( \mathcal{T}_n \) of \( \mathcal{EL} \) TBoxes and a fixed signature \( \Sigma \) such that

- the size of \( \mathcal{T}_n \) is upper-bounded by a polynomial in \( n \) and
- the size of the smallest uniform interpolant of \( \mathcal{T}_n \), w.r.t. \( \Sigma \) is lower-bounded by \( 2^{\Omega(2^{2n-1})} \).

Proof Sketch. For \( n \) a natural number, let the \( \mathcal{EL} \) TBox \( \mathcal{T}_n \) be given by

\[
\begin{align*}
A_1 \subseteq X_0 \cap \ldots \cap X_{n-1} \\
A_2 \subseteq X_0 \cap \ldots \cap X_{n-1} \\
\prod_{i \in \{1, \ldots, n\}} \exists r_i (X_i \cap X_0 \cap \ldots \cap X_{i-1}) \subseteq X_i & \quad i < n \\
\prod_{i \in \{1, \ldots, n\}} \exists r_i (X_i \cap X_0 \cap \ldots \cap X_{i-1}) \subseteq \neg X_i & \quad i < n \\
\prod_{i \in \{1, \ldots, n\}} \exists r_i (X_i \cap X_0 \cap \ldots \cap X_{i-1}) \subseteq X_i & \quad j < j < n \\
\prod_{i \in \{1, \ldots, n\}} \exists r_i (X_i \cap X_0 \cap \ldots \cap X_{i-1}) \subseteq B & \quad (34)
\end{align*}
\]

Obviously, the size of \( \mathcal{T}_n \) is polynomially bounded by \( n \). We now consider sets \( C_k \) of concept descriptions inductively defined by \( C_0 = \{A_1, A_2\} \) and \( C_{k+1} = \{ \exists r.C_{k} \mid \exists s.C_{2} \} \cup \{ C_1, C_2 \} \cap C_k \). We find that \( |C_{k+1}| = |C_k|^2 \) and consequently \( |C_k| \geq 2^{2^k} \). Thus, the set \( C_{2n-1} \) contains triply exponentially many different concepts, each of which is doubly exponential in the size of \( \mathcal{T}_n \).

Obviously, for any \( k \), every concept description from \( C_k \) uses only signature elements from \( A_1, A_2, r, s \).

It is rather straightforward to check that \( \mathcal{T}_n \models C \subseteq B \) holds for each concept \( C \in C_{2n-1} \); by induction on \( k \), we can show that for any \( C \in C_k \) with \( k < 2^n \) holds \( \mathcal{T}_n \models C \subseteq Y^n \cap \ldots \cap Y_{n-1} \) with

\[
Y_k^n = \begin{cases}
X_i & \text{if } \bar{k} \equiv 1 \mod 2 = 1 \\
X_i & \text{if } \bar{k} \equiv 0 \mod 2 = 0
\end{cases}
\]

i.e., \( Y_k^n \) indicates the \( k \)th bit of the number \( k \) in binary encoding.

Then, \( C \subseteq B \) follows via the last axiom of \( \mathcal{T}_n \).

Toward the claimed triple-exponential lower bound, we now show that every uniform interpolant of \( \mathcal{T}_n \) for \( \Sigma = \{A_1, A_2, B, r, s\} \) must contain for each \( C \in C_{2n-1} \) a GCI of the form \( C \subseteq B' \) with \( B' = B \) or \( B' = B \cap F \) for some \( F \) (where we consider structural variants –
i.e., concept expressions which are equivalent w.r.t. the empty knowledge base – as syntactically equal). Toward a contradiction, we assume that this is not the case, i.e., there is a uniform interpolant $T'$ and a $C \in C_{2n-1}$ where $C \subseteq B' \not\subseteq T'$ for any $B'$ containing $B$ as a conjunct.

Yet, since $C \subseteq B$ must be a consequence of $T'$, there must be a derivation of it. Looking at the derivation calculus from the last section, the last derivation step must be (ANDL) or (CUT). We can exclude (ANDL) since neither $\exists r.C' \subseteq B$ nor $\exists s.C' \subseteq B$ is the consequence of $T'$ for any $C' \in C_{2n-2}$ (which can be easily shown by providing appropriate witness models of $T'$). Consequently, the last derivation step must be an application of (CUT), i.e., there must be a concept $C \neq C$ such that $T' \models C \subseteq E$ and $T' \models E \subseteq B$. Without loss of generality, we assume that we consider a derivation where the branch of the derivation branch for $C \subseteq E$ has minimal depth.

We now distinguish two cases: either $E$ contains $B$ as a conjunct or not.

- First we assume $E = E' \cap B$, i.e. the CUT rule was used to derive $C \subseteq B$ from $C \subseteq E' \cap B$ and $E' \cap B \subseteq B$. The former cannot be contained in $T'$ by assumption, hence it must have been derived itself. Again, it cannot have been derived via (ANDL) for the same reasons as given above, which again leaves (CUT) as the only possible derivation rule for obtaining $C \subseteq E' \cap B$. Thus, there must be some concept $G$ with $T' \models C \subseteq G$ and $T' \models G \subseteq E' \cap B$. Once more, we distinguish two cases: either $G$ contains $B$ as a conjunct or not.

  - If $G$ contains $B$ as a conjunct, i.e., $G = G' \cap B$, the derivation of $C \subseteq E$ was not depth-minimal since there is a better proof where $C \subseteq B$ is derived from $C \subseteq G' \cap B$ and $G' \cap B \subseteq B$ via (CUT). Hence we have a contradiction.

  - If $G$ does not contain $B$ as a conjunct, the original derivation of $C \subseteq E$ was not depth-minimal since we can construct a better one that derives $C \subseteq B$ directly from $C \subseteq G$ and $G \subseteq B$ (the latter being derived from $G \subseteq E' \cap B$ via (ANDR)).

- Now assume $E$ does not contain $B$ as a conjunct. We construct $(\Delta, \tau^2)$, the “characteristic interpretation” of $C$ as follows ($\epsilon$ denoting the empty word):

  \[ \Delta = \{ w \mid w \in \{ r, s \}^*, \text{length}(w) < 2^n \} \]

  - We define an auxiliary function $\chi$ associating a concept expression to each domain element: we let $\chi(\epsilon) = C$ and for every $w, s \in \Delta$ with $\chi(w) = 3\forall r C_1 \exists s C_2$, we let $\chi(wr) = C_1$ and $\chi(ws) = C_2$.

  - The concepts and roles are interpreted as follows:

    \begin{itemize}
    \item $A^2_\epsilon = \{ w \mid \chi(w) = A_\epsilon \}$ for $\epsilon \in \{ 1, 2 \}$
    \item $B^2 = \{ \epsilon \}$
    \item $X^2_i = \{ w \mid \frac{\text{length}(w)}{2} \text{mod} \: 2 = 0 \}$ for $i < n$
    \item $\bar{X}^2_i = \{ w \mid \frac{\text{length}(w)}{2} \text{mod} \: 2 = 1 \}$ for $i < n$
    \item $\tau^2 = \{ (w, wr) \mid wr \in \Delta \}$
    \item $\alpha^2 = \{ (w, ws) \mid ws \in \Delta \}$
    \end{itemize}

  - It is straightforward to check that $\Delta$ is a model of $T_n$ and that $\epsilon \in C^2$. Consequently, due to our assumption, $\epsilon \in E^2$ must hold. Yet then, by construction, $E$ can only be a proper “structural superconcept” of $C$, i.e., $\emptyset \models C \subseteq E$ and $\emptyset \not\models E \subseteq C$ must hold. We now obtain $\hat{E}$ by enriching $E$ as follows: recursively, for every subexpression $G$ of $E$ satisfying $\emptyset \models G \subseteq C'$ for some $C' \in C_k$ for some $k < 2^n$, we substitute $G$ by $G \cap Y^k_0 \cap \ldots \cap Y^k_{n-1}$. Then, $\hat{E}$ directly corresponds to a finite tree interpretation $T'$ which is a model of $T_n$ (following from structural induction on subexpressions of $\hat{E}$) and the root individual of which satisfies $\hat{E}$ but not $C$ (by assumption). Yet, the root individual cannot satisfy any other concept expression $C''$ from $C_{2n-1} \setminus \{ C \}$ either, since this, via $\emptyset \models E \subseteq C''$, would imply $\emptyset \models C \subseteq C''$ which is not the case (by induction on $k$ one can show that there cannot be a homomorphism between the associated tree interpretations of any two distinct concepts from any $C_k$). In particular, we note that the root individual of $T'$ also does not satisfy $B$. Thus, we have found a model of $T_n$ witnessing $T_n \not\models E \subseteq B$, contradicting our assumption that $T' \models E \subseteq B$.

\[ \square \]

\section{Proof of Lemma 2}

Here, we prove a stronger version of Lemma 2 (the difference is the stronger statement [A4], which is used only within the induction-based proof of this lemma.)

Let $T$ be a normalized $\mathcal{EL}$ TBox and $C, D$ two $\mathcal{EL}$ concepts with $\text{sig}(C) \cup \text{sig}(D) \subseteq \text{sig}(T)$ such that $T \models C \subseteq D$. For any $A \in \text{sig}_G(T)$, let $\text{Pre}(A) = \{ M \subseteq \text{sig}_G(T) \mid T \models \bigwedge_{i \in M} B_i \subseteq A \}$. W.l.o.g., assume that

\[ C = \bigcap_{1 \leq j \leq n} A_j \cap \bigcap_{1 \leq k \leq m} \exists r_k . E_k \]

for $A_j \in \text{sig}_G(T)$ and $r_k \in \text{sig}_R(T)$. $E_k$ $\mathcal{EL}$ concepts with $\text{sig}(E_k) \subseteq \text{sig}(T)$ for $1 \leq k \leq m$. Then, for all conjunctions $D_k$ of $D$, the following is true: If $D_k \in \text{sig}_G(T)$, there is a set $M \in \text{Pre}(D_k) \subseteq \text{sig}_G(T)$ concepts such that for each element $B$ of $M$ holds at least one of the conditions [A1]-[A2]:

\begin{itemize}
  \item [(A1)] There is an $A_j$ in $C$ such that $A_j = B$.
  \item [(A2)] There are $r_k, E_k$ and there exists $B' \in \text{sig}_G(T)$ such that $T \models E_k \subseteq B' \subseteq B \exists r_k . B' \subseteq T$.
  \item [(A3)] There are $r_k, E_k$ such that $r_k = r'$ and $T \models E_k \subseteq D'$.
  \item [(A4)] There is $B \in \text{Pre}(C)$ such that $T \models B \subseteq \exists r . D' \text{ and } T \models \exists r . B' \subseteq T$.
\end{itemize}

\textbf{Proof.} We consider all rules, that could have been the last rule applied in order to obtain the above sequent and show by induction on the length of the proof that, in each case, the lemma holds. Rules $\mathbb{AXTOP}, \mathbb{AX}$ are the basecase, since each proof begins with one of them.

\((C \Rightarrow D \in T)\) In the case that $C \subseteq D \in T$ or $C \equiv D \in T$, the lemma holds due to the normalization. Axioms within $T$ can have the following form:

\begin{itemize}
  \item $C, D \in \text{sig}_G(T)$. In this case, $\{ C \} \in \text{Pre}(D)$. Therefore, condition [A1] holds.
  \item $C \in \text{sig}_G(T), D = D_1 \cap \ldots \cap D_m$ with $D_1, \ldots, D_m \in \text{sig}_G(T)$. In this case, for each $D_i$ with $1 \leq i \leq m$ holds $\{ C \} \in \text{Pre}(D_i)$. Therefore, condition [A1] holds for each $D_i$.\end{itemize}
• $C \in \text{sig}_C(T), D = \exists r'.D'$ with $D' \in \text{sig}_C(T)$. This case corresponds to the condition [A4].

(AXTop) Since the conjunction is empty in case $D = \top$, the lemma holds.

(AX) Since $C = D$, for each $D_i$ there is a conjunct $C_i$ of $C$ with $C_i = D_i$. If $D_i \in \text{sig}_C(T)$, condition [A1] of the lemma holds. Otherwise, [A3].

(EX) If $E$ was the last applied rule, then $D_i = \exists r_k.D'$ and $T \vdash D_k \subseteq D'$. Therefore, [A3] of the lemma holds.

(ANDL) Assume that $C' \cap C'' = C$ such that $C' \subseteq D$ is the antecedent. By induction hypothesis, the lemma holds for $C' \subseteq D$. Since all conjuncts of $C''$ are also conjuncts of $C$, the lemma holds also for $C \subseteq D$.

(ANDR) Assume that $D = D_1 \cap D_2$, therefore, $C \subseteq D_1$ and $C \subseteq D_2$ is the antecedent. By induction hypothesis, the lemma holds for both elements of the antecedent, $C \subseteq D_1$ and $C \subseteq D_2$. Since all conjuncts of $C''$ are also conjuncts of $C$, the lemma holds also for $C \subseteq D$.

(CUT) By induction hypothesis, the lemma holds for both elements of the antecedent, $C \subseteq C_1$ and $C \subseteq C_2$. W.l.o.g., assume that $C = \prod_{1 \leq i \leq s} A_i$. Then, let $M_{\text{new}} = M_1 \cup M_{\text{old}}$.

(A3) There are $r_k, E_k$ such that $r_k = r'$ and $T \vdash E_k \subseteq E'$. Then [A2] holds for $C \subseteq B_1$, since $T \vdash E_k \subseteq B$ and $B = \exists r_k B'$. Therefore, $M_{\text{new}}$ holds.

(A4) There is $B'' \in \text{nt} \text{ such that } T \vdash B'' \subseteq \exists r'.E', T \vdash C \subseteq B''$ and there is a set $M'' \in \text{Pre}(B'')$ such that for each element $B'$ of $M''$ holds at least one of the conditions [A1]-[A2] w.r.t. $C \subseteq B'$. Let $M_{\text{new}}(B_i) = M_i$ and $M_{\text{old}} \subseteq M_i$ be the set of all such $B_i$. Then, let $M_{\text{new}} = M_1 \cup \{M_{\text{new}}(B_i) \mid B_i \in M_{\text{old}}\}$.

Clearly, $M_{\text{new}} \in \text{Pre}(D_i)$ and [A1] or [A2] holds for each $B_i \in M_{\text{new}}$ w.r.t. $C \subseteq B_1$, i.e., the lemma holds for $C \subseteq D_i$.


(A3) There are $r_k, E_k$ such that $r_k = r'$ and $T \vdash E_k \subseteq D'$. Then, for $C \subseteq \exists r_k E_k$ one of [A3], [A4] holds.

(A4) There is $B'' \in \text{nt}$ such that $T \vdash B'' \subseteq \exists r'.E', T \vdash C \subseteq B''$ and there is a set $M'' \in \text{Pre}(B'')$ of sig_c(T) concepts such that for each element $B'$ of $M''$ holds at least one of the conditions [A1]-[A2] w.r.t. $C \subseteq B'$.

Since $T \vdash B'' \subseteq D_i$, [A4] holds for $T \vdash C \subseteq D_i$.

(A4) There is $B \in \text{nt}$ such that $T \vdash B \subseteq \exists r'.D', T \vdash C \subseteq B$ and there is a set $M' \in \text{Pre}(B)$ such that for each element $B'$ of $M$ holds at least one of the conditions [A1]-[A2] w.r.t. $C \subseteq B'$. Then, we have the same situation as above with two subsumptions $C \subseteq C_1$ and $C_1 \subseteq B$, where $B \in \text{sig}_C(T)$. Therefore, the argumentation is the same as above implying that the claim of the lemma holds for $C \subseteq B$, i.e., there is $M_i \in \text{Pre}(B)$ such that [A1] or [A2] holds for each $B_i \in M_i$. Then, [A4] holds for $C \subseteq D_i$.

D Proofs for Section 5

Theorem 2

Let $T$ be a normalized $\mathcal{E}$L TBox, $\Sigma$ a signature and $A \in \text{sig}_C(T)$.

1. For each $t \in L(G^2(T, \Sigma, A))$, there is a concept with $tC = t$ and $\text{sig}(C) \subseteq \Sigma$ such that $T \vdash C \subseteq A$.

2. For each $t \in L(G^2(T, \Sigma, A))$, there is a concept $C$ with $tC = t$ and $\text{sig}(C) \subseteq \Sigma$ such that $T \vdash A \subseteq C$.

Proof. It is easy to check in Definition 2 that the grammars derive only terms containing atomic concepts and roles from $\Sigma$, since $n_{\mathcal{E}} \rightarrow B$ only if $B \in \Sigma$ and $n_{\mathcal{E}} \rightarrow \exists r(t)$ only if $r \in \Sigma$. For any $A \in \text{sig}_C(T)$ and any $t \in L(G^2(T, \Sigma, A))$, we have the same situation as above with two subsumptions $C \subseteq C_1$ and $C_1 \subseteq B$, where $B \in \text{sig}_C(T)$. Therefore, the argumentation is the same as above implying that the claim of the lemma holds for $C \subseteq B$, i.e., there is $M_i \in \text{Pre}(B)$ such that [A1] or [A2] holds for each $B_i \in M_i$. Then, [A4] holds for $C \subseteq D_i$.
2. The proof of soundness of $G^C(T, \Sigma)$ can be done in the same manner. Let $t$ be a term such that $t \in L(G^C(T, \Sigma, A))$. We prove the theorem by induction on the maximal nesting depth of functions in $t$.

- Assume that $t$ is an atomic concept $B$. $B$ can only be derived from $n_1 \in \Sigma$ by empty transitions (GR3) and, since $n_1$ is $\lambda$-free, the rule (GR2). Let $B_1, \ldots, B_n$ be such that $n_1 \rightarrow n_{B_1} \rightarrow \ldots \rightarrow n_{B_n} \rightarrow n_2$. Then, by Definition 4 for each pair $B_i, B_{i+1}$ holds $T \models B_i$, $B_{i+1}$. For $B_{i+1}$, $B$ holds $T \models B_n \in \Delta$ and for $A, B_1$ holds $T \models A \in B_1$. It follows that also $T \models A \in \Sigma$ with $t = t_{n_2}$.

- Assume that $t = \exists r(t')$ for some term $t'$. Then, the derivation of $t$ from $n_1$ starts with $n_1$ empty transitions (GR3) such that $n_2$ for some $B' \in \Sigma(T)$ is reached, and a subsequent application of a non-empty transition (GR3) such that $n_3$ for some $B \in \Sigma(T)$ is reached. As argued above about the applications of empty transitions, $T \models A \subseteq B'$ holds. Moreover, by Definition 4 holds $T \models B' \subseteq B$. Therefore, $T \models A \subseteq B$. Let $C'$ be a concept with $t = t_{n_3}$. By induction hypothesis, $T \models B \subseteq C'$. Therefore, $T \models A \subseteq \exists r(C')$ with $t = t_{n_3} C'$.

- Assume that $t = \bigcap (t_1, \ldots, t_n)$ for a set of terms $t_1, \ldots, t_n$. Then, the derivation of $t$ from $n_1$ starts with $n_1$ empty transitions (GR3) such that $n_2$ for some $B' \in \Sigma(T)$ is reached, and a subsequent application of Definition 4 such that, for a set of concepts $B_i \in \Sigma(T)$ with $1 \leq i \leq n$ and $t_i \in L(G^C(T, \Sigma, n_1))$, $n_3$ is reached. As argued above about the applications of empty transitions, $T \models A \subseteq B'$ holds. Let $C'$ be a concept with $t = t_{n_3}$. By induction hypothesis, $T \models B_1 \subseteq C'$. By Definition 4 holds $T \models B' \subseteq C'$. Therefore, $T \models A \subseteq C' \cap \ldots \cap C''$, while $t = t_{n_3} C' \cap \ldots \cap C''$. \(\Box\)

We start the proof of completeness of a Lemma.

**Lemma 5** Let $T$ be a normalized $\mathcal{EL}$ TBox, $A \in \Sigma_{G^C(T)}$ and $\nu \in \Sigma_{G^C(T)}$. Let $C$ be an $\mathcal{EL}$ concept such that $T \models A \subseteq \exists r(C)$. Then, there are $B_1, B_2 \in \Sigma_{G^C(T)}$ with $B_1 \equiv \exists r B_2$ in $T$ such that $T \models A \subseteq B_1$, $T \models B_2 \subseteq C$.

**Proof.** Lemma 16 in [10] states that for a general $\mathcal{EL}$ TBox $T$ with $T \models C_1 \subseteq \exists r C_2$, where $C_1, C_2$ are $\mathcal{EL}$-concepts one of the following holds:

- there is a conjunct $\exists r C'$ of $C_1$ such that $T \models C' \subseteq C_2$;
- there is a subconcept $\exists r C'$ of $T$ such that $T \models C_1 \subseteq \exists r C'$ and $T \models C' \subseteq C_2$.

The first condition does not hold in this lemma, since $A \in \Sigma_{G^C(T)}$. Moreover, since in our case $T$ is normalized, for each subconcept $\exists r C'$ of $T$ containing an existential restriction holds; there is an atomic concept $B_2 \in \Sigma_{G^C(T)}$ such that $B_2 = C'$ and there is an axiom of the form $B_1 \equiv \exists r B_2$ in $T$ with $B_1 \in \Sigma_{G^C(T)}$. Additionally, from the above $\Sigma_{G^C(T)}$ holds $T \models A \subseteq \exists r B_2$ and $T \models B_2 \subseteq C$. Since $T \models B_1 \equiv \exists r B_2$, it follows that also $T \models A \subseteq B_1$. \(\Box\)

We proceed with the proving the two parts of Theorem 3. In what follows, we say that a concept $C$ can be obtained from a concept $C'$ by weakening, meaning that $C$ can obtained from $C'$ by adding arbitrary conjuncts to arbitrary subexpressions.

**Theorem** Let $T$ be a normalized $\mathcal{EL}$ TBox, $\Sigma$ a signature and $A \in \Sigma_{G^C(T)}$. W.l.o.g., we can assume that there is a concept $C$ with

$$C = \bigcap_{1 \leq j \leq n} A_j \cap \bigcap_{1 \leq k \leq m} \exists r_k E_k$$

with $A_j \in \Sigma$ for $1 \leq j \leq n$, $r_k \in \Sigma$ for $1 \leq k \leq m$ and $E_k$ with $1 \leq k \leq m$ a set of $\mathcal{EL}$ concepts such that $\Sigma(E_k) \subseteq \Sigma$. Further, w.l.o.g., we can assume that all $A_j$ are pairwise different.

1. We show that, for each such general $C$ with $\Sigma(C) \subseteq \Sigma$ and $T \models C \subseteq A$, there is a concept $C'$ such that $C$ can be obtained from $C'$ by weakening and $t_{C'} \in L(G^C(T, \Sigma, A))$. We prove the claim by induction of the role depth of $C$.

- Assume role depth is $0$. Then $C$ is a conjunction of atomic concepts, i.e., $m = 0$ and $C = \bigcap_{1 \leq j \leq n} A_j$. Then, by Lemma 5 there is a set $M' \in \mathcal{P}(A)$ (of atomic concepts such that, for each $B \in M'$, there is an $A_j$ with $A_j = B$. Therefore, each $B \in M'$ is in $\Sigma$. Let $C_1 = \bigcap_{B \in M'} B$. Since $M' \subseteq \{A_1, \ldots, A_n\}$, $C$ can be obtained from $C_1$ by weakening. By Definition 8 there is a rule $n_4 \rightarrow n_{\{n_B, \ldots, n_B\}}$ with $\{B_1, \ldots, B_n\} = M'$. Since each $B \in M'$ is in $\Sigma$, we obtain by (GL5) $n_4 \rightarrow B$. Since our grammars operate on unordered trees, it follows that $n_4 \rightarrow G^C(T, \Sigma, A)$.

2. Assume the role depth is greater than $0$. As in the case above, there is a set $M' \in \mathcal{P}(A)$ (of atomic concepts such that, for each $B \in M'$, [A1] or [A2] holds. Let $M' = \bigcap_{1 \leq j \leq n} \{A_j, \ldots, A_n\}$ and $M'' = M' \setminus M$. Let $C_1 = \bigcap_{B \in M''} B$, and $C_2 = \bigcap_{1 \leq j \leq n} \exists r_i E_i$ with $\{\exists r_i E_i, \ldots, \exists r_i E_i\} = \{\exists r E \subseteq \} for one of $B_1 \in M''$ [A2] such that there exists $B' \in \Sigma(T)$ with $T \models E \subseteq B'$ and $B \equiv B' \subseteq T$. Clearly, $C$ can be obtained from $C_1 \cap C_2$ by weakening. By Definition 8 there is a rule $n_4 \rightarrow n_{\{n_B, \ldots, n_B\}}$ with $\{B_1, \ldots, B_n\} = M'$. Moreover, for all $B \in M''$ holds $n_B \rightarrow B$ and for all $B_1 \in M''$, there is $\exists r_i E_i$ such that there exists $B' \in \Sigma(T)$ with $T \models E_i \subseteq B'$ and $B_i \equiv \exists r_i E_i \subseteq T$. By Definition 8 (GL8), $n_B \rightarrow D_i (n_B)$.

By induction hypothesis, there is a concept $E_i$ such that $n_B \rightarrow \bigcap_{T \models C'} \exists r_i E_i$ and $E_i$ can be obtained from $E_i$ by weakening. Therefore, $n_B \rightarrow \bigcap_{T \models C'} \exists r_i E_i$ and $E_i$ can be obtained from $\exists r_i E_i$ by weakening. Let $C''' = C_1 \cap \bigcap_{B \in M'} \exists r_i E_i$. Then, $C$ can be obtained from $C'''$ by weakening. Since our grammars operate on unordered trees, we obtain $n_4 \rightarrow G^C(T, \Sigma, A)$.
2. We proceed with showing that for each such general \( C \) with \( \text{sig}(C) \subseteq \Sigma \) and \( \mathcal{T} \models A \subseteq C \) holds: \( t_C \in L(G^\subseteq(\mathcal{T}, \Sigma, A)) \). We prove the claim by induction of the role depth of \( C \). For each \( A_j \), we know that \( \mathcal{T} \models A \subseteq A_j \) and \( A_j \in \Sigma \), i.e., \( A_j \in \text{Post}_{\text{Base}}(A) \).

By Definition 2, \( n_{A_j} \rightarrow A_j \) for all \( A_j \). By Definition 4, \( n_A \rightarrow \cap(n_{A_1}, \ldots, n_{A_k}) \), and, therefore, \( t_C \in L(G^\subseteq(\mathcal{T}, \Sigma, A)) \).

Assume a role depth > 0. For each \( \exists r_k.E_k \), it follows from Lemma 5 that there are \( B_1, B_2 \in \text{sig}_C(\mathcal{T}) \) with \( B_1 \equiv \exists r_k.B_2 \in \mathcal{T} \) such that \( \mathcal{T} \models A \subseteq B_1, \mathcal{T} \models B_2 \subseteq E_k \). Since \( r_k \in \Sigma \), follows that \( \exists r_k.B_2 \in \text{Post}_{\text{Base}}(A) \). Moreover, by induction hypothesis follows that \( t_{E_k} \in L(G^\subseteq(\mathcal{T}, \Sigma, B_2)) \). An application of (GR3) in combination with Definition 4 yields \( t_C \in L(G^\subseteq(\mathcal{T}, \Sigma, A)) \). \( \square \)