Computing the lcs w.r.t. General $\mathcal{EL}^+$-TBoxes

Anni-Yasmin Turhan* and Benjamin Zarrieß
{turhan,zarriess}@tcs.inf.tu-dresden.de

Chair for Automata Theory
Institute for Theoretical Computer Science
Technische Universität Dresden, Germany

Abstract. Recently, exact conditions for the existence of the least common subsumer (lcs) computed w.r.t. general $\mathcal{EL}$-TBoxes have been devised [13]. This paper extends these results and provides necessary and suffcient conditions for the existence of the lcs w.r.t. $\mathcal{EL}^+$-TBoxes. We show decidability of the existence in PTime and polynomial bounds on the maximal role-depth of the lcs, which in turn yields a computation algorithm for the lcs w.r.t. $\mathcal{EL}^+$-TBoxes.

1 Introduction

In the area of Description Logics (DLs) the least common subsumer (lcs) is an inference that is applied to a collection of concepts and yields a complex concept that captures all commonalities of the input concepts. Unfortunately, the lcs doesn't need to exist, if computed w.r.t. general $\mathcal{EL}$-TBoxes [3]. Let's consider the TBox axioms:

$$\begin{aligned}\mathsf{Woman} &\sqsubseteq \mathsf{Human} \sqcap \exists \mathsf{has\text{-}Grandparent.Woman},\\ \mathsf{Man} &\sqsubseteq \mathsf{Human} \sqcap \exists \mathsf{has\text{-}Grandparent.Man},\\ \mathsf{Human} &\sqsubseteq \exists \mathsf{has\text{-}Parent.Human}.\end{aligned}$$

We want to compute the lcs of Woman and Man. Both are Human and have Grandparents that are Woman or Man, respectively. This leads to a cyclic definition and thus the *least* common subsumer cannot be captured by a finite $\mathcal{EL}$-concept, since this would need to express the cycle.

The DL $\mathcal{EL}^+$ allows, in addition to $\mathcal{EL}$, for role inclusion axioms as for example:

$$\mathsf{has\text{-}Parent} \circ \mathsf{has\text{-}Parent} \sqsubseteq \mathsf{has\text{-}Grandparent}$$

If we consider the lcs of Woman and Man w.r.t. a knowledge base that in addition to the axioms above also contains this role inclusion axiom, then the lcs exists and is just Human. We can observe that the existence of the lcs does not merely depend on whether the TBox is cyclic.

* Partially supported by the German Research Foundation (DFG) in the Collaborative Research Center 912 "Highly Adaptive Energy-Efficient Computing".

In fact, for cyclic definitorial $\mathcal{EL}$-TBoxes exact conditions for the existence of the lcs have been devised [4]. For the case of general $\mathcal{EL}$-TBoxes such conditions were shown recently in [13]. In this paper we want to extend these results to $\mathcal{EL}^+$. Since in $\mathcal{EL}^+$ no expressive power on the level of concepts is added compared to $\mathcal{EL}$ and every general $\mathcal{EL}$-TBox is also a general $\mathcal{EL}^+$-TBox, the non-existence of the lcs w.r.t. to general TBoxes carries over to $\mathcal{EL}^+$.

There are also several approaches to compute the lcs even in the presence of general TBoxes. In [9] an extension of $\mathcal{EL}$ with greatest fixpoints was introduced, where the lcs concepts always exist. Computation algorithms for approximate solutions for the lcs were devised in [6, 11] and were also extended to $\mathcal{EL}^+$ [8]. The last two methods simply compute the lcs-concept up to a given k, a bound on the maximal nestings of quantifiers. If the lcs exists and a large enough k was given, then these methods yield the exact solutions. However, to obtain the *least* common subsumer by these methods in practice, a decision procedure for the existence of the lcs w.r.t. general $\mathcal{EL}^+$-TBoxes and a method for computing a sufficient bound k is still needed. This paper provides this method for the lcs. Similar to the case of approximate solutions for the lcs [8, 11], it also turns out in our case that the results shown for $\mathcal{EL}$ (in [13]) can be easily extended to $\mathcal{EL}^+$.

In this paper we first introduce basic notions for the DL $\mathcal{EL}^+$ and its canonical models, which serve as a basis for the characterization of the lcs in the subsequent sections. There we show that the characterization can be used to verify whether a given common subsumer is indeed the least one and show that the size of the lcs, if it exists, is polynomially bounded in the size of the input. These results yield a decision procedure for the existence problem for the lcs in $\mathcal{EL}^+$. The paper ends with some conclusions.

2 Preliminaries

2.1 The Description Logic $\mathcal{EL}^+$

Let $\mathsf{N_C}$ and $\mathsf{N_R}$ be disjoint sets of *concept* and *role names*. Let $A \in \mathsf{N_C}$ and $r \in \mathsf{N_R}$. *$\mathcal{EL}$-concepts* are built according to the syntax rule

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

An *interpretation* $\mathcal{I} = (\Delta^\mathcal{I}, \cdot^\mathcal{I})$ consists of a non-empty domain $\Delta^\mathcal{I}$ and a function $\cdot^\mathcal{I}$ that assigns subsets of $\Delta^\mathcal{I}$ to concept names and binary relations on $\Delta^\mathcal{I}$ to role names. The function is extended to complex concepts in the usual way. For a detailed description of the semantics of DLs see [1].

Let C, D denote $\mathcal{EL}$-concepts. A *general concept inclusion* (GCI) is an expression of the form $C \sqsubseteq D$. The DL $\mathcal{EL}^+$ allows also for *role inclusion axioms* (RIAs) of the form $r_1 \circ ... \circ r_n \sqsubseteq r$ with $n \geq 1$ and $r_i \in \mathsf{N_R}$ for $i = 1, ..., n$ and $r \in \mathsf{N_R}$. A *(general) TBox* $\mathcal{T}$ is a finite set of GCIs and an RBox $\mathcal{R}$ is a finite set of RIAs. An *ontology* $\mathcal{O}$ consists of a TBox $\mathcal{T}$ and an RBox $\mathcal{R}$, denoted by $\mathcal{O} = (\mathcal{T}, \mathcal{R})$.[1]

[1] Since we only use the DL $\mathcal{EL}^+$, we write 'concept' instead of '$\mathcal{EL}$-concept' and assume all TBoxes and RBoxes to be written in $\mathcal{EL}^+$ in the following.

A GCI $C \sqsubseteq D$ is satisfied in an interpretation $\mathcal{I}$ iff $C^\mathcal{I} \subseteq D^\mathcal{I}$. The $\circ$ operator on roles is interpreted as

$$r^\mathcal{I} \circ s^\mathcal{I} := \{(d, f) \mid \exists e \in \Delta^\mathcal{I} \text{ s.t. } (d, e) \in r^\mathcal{I} \wedge (e, f) \in s^\mathcal{I}\}.$$

A RIA $r_1 \circ ... \circ r_n \sqsubseteq r$ is satisfied in an interpretation $\mathcal{I}$ if $r_1^\mathcal{I} \circ ... \circ r_n^\mathcal{I} \subseteq r^\mathcal{I}$. An interpretation $\mathcal{I}$ is a *model* of $\mathcal{T}$, if it satisfies all GCIs in $\mathcal{T}$. An interpretation $\mathcal{I}$ is a *model* of $\mathcal{R}$ if it satisfies all RIAs in $\mathcal{R}$. $\mathcal{I}$ is a model of an ontology $\mathcal{O} = (\mathcal{T}, \mathcal{R})$ if $\mathcal{I}$ is a model of $\mathcal{T}$ and $\mathcal{R}$.

An important reasoning task considered for DLs w.r.t. TBoxes and RBoxes is *subsumption*. A concept C *is subsumed by* a concept D w.r.t. an ontology $\mathcal{O}$ (denoted by $C \sqsubseteq_\mathcal{O} D$) iff $C^\mathcal{I} \subseteq D^\mathcal{I}$ holds in all models $\mathcal{I}$ of $\mathcal{O}$. A concept C is *equivalent* to a concept D w.r.t. $\mathcal{O}$ (denoted by $C \equiv_\mathcal{O} D$) iff $C \sqsubseteq_\mathcal{O} D$ and $D \sqsubseteq_\mathcal{O} C$ hold. Similarly, a concept C is subsumed by a concept D w.r.t. an RBox $\mathcal{R}$ (denoted by $C \sqsubseteq_\mathcal{R} D$) iff $C^\mathcal{I} \subseteq D^\mathcal{I}$ holds in all models $\mathcal{I}$ of $\mathcal{R}$.

Subsumption w.r.t. ontologies can be decided for $\mathcal{EL}^+$ in polynomial time [5]. Based on subsumption the inference we are interested in, the *least common subsumer* (lcs) is defined.

Definition 1 (least common subsumer). *Let C, D be concepts and $\mathcal{O}$ an ontology. The concept E is the* lcs of C, D w.r.t. $\mathcal{O}$ *(lcs$_\mathcal{O}(C, D)$) if the properties*

1. *$C \sqsubseteq_\mathcal{O} E$ and $D \sqsubseteq_\mathcal{O} E$, and*
2. *$C \sqsubseteq_\mathcal{O} F$ and $D \sqsubseteq_\mathcal{O} F$ implies $E \sqsubseteq_\mathcal{O} F$.*

are satisfied. If a concept E satisfies Property 1 it is a common subsumer of C and D w.r.t. $\mathcal{O}$.

The lcs in a DL that offers conjunction is unique up to equivalence, thus we speak of *the* lcs. In contrast to this, common subsumers are not unique, thus we write $F \in \text{cs}_\mathcal{O}(C, D)$.

The *role depth* ($rd(C)$) of a concept C denotes the maximal number of nestings of $\exists$ in C. If in Definition 1 the concepts E and F have a role-depth bound less than or equal k, then E is the *role-depth bounded lcs* (k-lcs$_\mathcal{O}(C, D)$) of C and D w.r.t. $\mathcal{O}$, which is also unique up to equivalence.

2.2 Canonical Models and Simulation Relations

The correctness proof of the computation algorithms for the lcs depends on the characterization of subsumption. In case of the lcs in $\mathcal{EL}$ without a TBox, homomorphisms between syntax trees of concepts [2] were used. A characterization w.r.t. general $\mathcal{EL}$-TBoxes using *canonical models* and *simulations* was given in [10]. This characterization was extended to $\mathcal{EL}^+$ and RBoxes in [12].

Let X be a concept, a TBox, an RBox or an ontology, then $\mathsf{sub}(X)$ denotes the subconcepts in X.

Definition 2 (canonical model [12]). *Let C be a concept and $\mathcal{O} = (\mathcal{T}, \mathcal{R})$ an ontology. The canonical model $\mathcal{I}_{C,\mathcal{O}}$ of C and $\mathcal{O}$ is defined as follows:*

- $\Delta^{\mathcal{I}_{C,\mathcal{O}}} := \{d_C\} \cup \{d_{C'} \mid \exists r.C' \in \mathsf{sub}(C) \cup \mathsf{sub}(\mathcal{T})\}$;
- $A^{\mathcal{I}_{C,\mathcal{O}}} := \{d_D \mid D \sqsubseteq_{\mathcal{O}} A\}$, *for all* $A \in \mathsf{N_C}$;
- $r^{\mathcal{I}_{C,\mathcal{O}}} := \{(d_D, d_{D'}) \mid D \sqsubseteq_{\mathcal{O}} \exists r.D'$ *for* $D' \in \mathsf{sub}(\mathcal{T})$ *or*
  $D \sqsubseteq_{\mathcal{R}} \exists r.D'$ *for* $D' \in \mathsf{sub}(C)\}$ *for all* $r \in \mathsf{N_R}$.

To identify properties of canonical models we use *simulation relations* between interpretations.

Definition 3 (simulation). *Let $\mathcal{I}_1$ and $\mathcal{I}_2$ be interpretations. $\mathcal{S} \subseteq \Delta^{\mathcal{I}_1} \times \Delta^{\mathcal{I}_2}$ is called a* simulation *from $\mathcal{I}_1$ to $\mathcal{I}_2$ if the following conditions are satisfied:*

1. *For all concept names $A \in \mathsf{N_C}$ and all $(e_1, e_2) \in \mathcal{S}$ it holds: $e_1 \in A^{\mathcal{I}_1}$ implies $e_2 \in A^{\mathcal{I}_2}$.*
2. *For all role names $r \in \mathsf{N_R}$ and all $(e_1, e_2) \in \mathcal{S}$ and all $f_1 \in \Delta^{\mathcal{I}_1}$ with $(e_1, f_1) \in r^{\mathcal{I}_1}$ there exists $f_2 \in \Delta^{\mathcal{I}_2}$ such that $(e_2, f_2) \in r^{\mathcal{I}_2}$ and $(f_1, f_2) \in \mathcal{S}$.*

To denote an interpretation $\mathcal{I}$ together with a particular element $d \in \Delta^{\mathcal{I}}$ we write $(\mathcal{I}, d)$. It holds that $(\mathcal{I}, d)$ *is simulated by* $(\mathcal{J}, e)$ (written as $(\mathcal{I}, d) \lesssim (\mathcal{J}, e)$) if there exists a simulation $\mathcal{S} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{J}}$ with $(d, e) \in \mathcal{S}$. The relation $\lesssim$ is a preorder, i.e. it is reflexive and transitive. $(\mathcal{I}, d)$ is *simulation-equivalent* to $(\mathcal{J}, e)$ (written as $(\mathcal{I}, d) \simeq (\mathcal{J}, e)$) if $(\mathcal{I}, d) \lesssim (\mathcal{J}, e)$ and $(\mathcal{J}, e) \lesssim (\mathcal{I}, d)$ holds.

Now we summarize some important properties of canonical models in $\mathcal{EL}^+$ that were shown in [12].

Lemma 1. *Let C be a concept and $\mathcal{O}$ an ontology.*

1. *$\mathcal{I}_{C,\mathcal{O}}$ is a model of $\mathcal{O}$.*
2. *For all models $\mathcal{I}$ of $\mathcal{O}$ and all $d \in \Delta^{\mathcal{I}}$ holds:*
   $d \in C^{\mathcal{I}}$ iff $(\mathcal{I}_{C,\mathcal{O}}, d_C) \lesssim (\mathcal{I}, d)$.
3. *$C \sqsubseteq_{\mathcal{O}} D$ iff $d_C \in D^{\mathcal{I}_{C,\mathcal{O}}}$ iff $(\mathcal{I}_{D,\mathcal{O}}, d_D) \lesssim (\mathcal{I}_{C,\mathcal{O}}, d_C)$.*

This lemma gives us a characterization of subsumption by means of canonical models. In order to employ it for the characterization of the lcs, we need to recall some operations on interpretations.

Starting from an element of the domain of an interpretation as the root, the interpretation can be unraveled into a possibly infinite tree. The nodes of the tree are words that correspond to paths starting in d. The word $\pi = dr_1d_1r_2d_2r_3...$ is a path in an interpretation $\mathcal{I}$ if the domain elements d_i and d_{i+1} are connected via $r_{i+1}^{\mathcal{I}}$ for all i.

Definition 4 (tree unraveling of an interpretation). *Let $\mathcal{I}$ be an interpretation w.r.t. names from $\mathsf{N_C}$ and $\mathsf{N_R}$ with $d \in \Delta^{\mathcal{I}}$. The tree unraveling $\mathcal{I}_d$ of $\mathcal{I}$ in d is defined as follows:*

$$\Delta^{\mathcal{I}_d} := \{dr_1d_1r_2...r_nd_n \mid (d_i, d_{i+1}) \in r_{i+1}^{\mathcal{I}} \wedge 0 \leq i < n \wedge d_0 = d\};$$
$$A^{\mathcal{I}_d} := \{\sigma d' \mid \sigma d' \in \Delta^{\mathcal{I}_d} \wedge d' \in A^{\mathcal{I}}\},\ \textit{for all } A \in \mathsf{N_C};$$
$$r^{\mathcal{I}_d} := \{(\sigma, \sigma r d') \mid (\sigma, \sigma r d') \in \Delta^{\mathcal{I}_d} \times \Delta^{\mathcal{I}_d}\},\ \textit{for all } r \in \mathsf{N_R}.$$

The *length* of an element $\sigma \in \Delta^{\mathcal{I}_d}$ (denoted by $|\sigma|$), is the number of role names occurring in σ. If σ is of the form $dr_1d_1r_2...r_md_m$, then d_m is the *tail* of σ denoted by $\mathsf{tail}(\sigma) = d_m$. The interpretation $\mathcal{I}_d^\ell$ denotes the finite subtree rooted in d of the tree unraveling $\mathcal{I}_d$ containing all elements up to depth ℓ. Such a tree can be translated into an *ℓ-characteristic concept* of an interpretation $(\mathcal{I}, d)$.

Definition 5 (characteristic concept). *Let $(\mathcal{I}, d)$ be an interpretation. The* ℓ-characteristic concept $X^\ell(\mathcal{I}, d)$ *is defined as follows:* [2]

$$X^0(\mathcal{I}, d) := \bigsqcap\{A \in \mathsf{N_C} \mid d \in A^{\mathcal{I}}\}$$
$$X^\ell(\mathcal{I}, d) := X^0(\mathcal{I}, d) \sqcap \bigsqcap_{r \in \mathsf{N_R}} \bigsqcap\{\exists r.X^{\ell-1}(\mathcal{I}, d') \mid (d, d') \in r^{\mathcal{I}}\}$$

Another operation that we will use later is the product of two interpretations that is defined as follows.

Definition 6 (product of interpretations). *Let $\mathcal{I}$ and $\mathcal{J}$ be interpretations. The* product interpretation $\mathcal{I} \times \mathcal{J}$ *is defined by*

$$\Delta^{\mathcal{I}\times\mathcal{J}} := \Delta^{\mathcal{I}} \times \Delta^{\mathcal{J}};$$
$$A^{\mathcal{I}\times\mathcal{J}} := \{(d, e) \mid (d, e) \in \Delta^{\mathcal{I}\times\mathcal{J}} \wedge d \in A^{\mathcal{I}} \wedge e \in A^{\mathcal{J}}\}, \textit{ for all } A \in \mathsf{N_C};$$
$$r^{\mathcal{I}\times\mathcal{J}} := \{((d, e), (f, g)) \mid ((d, e), (f, g)) \in \Delta^{\mathcal{I}\times\mathcal{J}} \times \Delta^{\mathcal{I}\times\mathcal{J}} \wedge (d, f) \in r^{\mathcal{I}} \wedge (e, g) \in r^{\mathcal{J}}\}, \textit{ for all } r \in \mathsf{N_R}.$$

3 Existence of Least Common Subsumers

In this section we develop a decision procedure for the problem whether for two given concepts and a given ontology the least common subsumer of these two concepts w.r.t. the given ontology exists. If not stated otherwise, the two input concepts are denoted by C and D and the ontology by $\mathcal{O}$.

We follow the method used in [13], which is based on operations and relations on canonical models. The only difference compared to the setting in [13] is the presence of RIAs in the RBox. Fortunately, canonical models of concepts w.r.t. GCIs and RIAs with the properties given in Lemma 1 can also be obtained as in the case where we have only GCIs. Therefore, the results shown for $\mathcal{EL}$ in [13] can be easily adopted to the case of $\mathcal{EL}^+$.

Similar to the approach used in [4] we proceed by the following steps:

1. Devise a method to identify lcs-candidates for the lcs. The set of lcs-candidates is a possibly infinite set of common subsumers of C and D w.r.t. $\mathcal{O}$, such that if the lcs exists then one of these lcs-candidates actually is the lcs.

2. Characterize the existence of the lcs. Find a condition such that the problem whether a given common subsumer of C and D w.r.t. $\mathcal{O}$ is least (w.r.t. $\sqsubseteq_{\mathcal{O}}$), can be decided by testing this condition.

[2] For a set M of concepts we write $\bigsqcap M$ as shorthand for $\bigsqcap_{F\in M} F$. If M is empty, then $\bigsqcap M$ is equal to $\top$.

3. Establish an upper bound on the role-depth of the lcs. We give a bound ℓ such that *if the lcs exists*, then it has a role-depth less than or equal ℓ. By the use of such an upper bound one needs to check only for finitely many of the lcs-candidates if they are least (w.r.t. $\sqsubseteq_{\mathcal{O}}$).

The next subsection addresses the first two problems, afterwards we show that such a desired upper bound exists.

3.1 Characterizing the Existence of the lcs

The characterization presented here is based on the product of canonical models. This product construction is adopted from [3, 9] where it was used to compute the lcs in $\mathcal{EL}$ with gfp-semantics and in the DL $\mathcal{EL}^{\nu}$, respectively.

To obtain the k-lcs$_{\mathcal{O}}(C, D)$ we build the product of the canonical models $(\mathcal{I}_{C,\mathcal{O}}, d_C)$ and $(\mathcal{I}_{D,\mathcal{O}}, d_D)$ and then take the k-characteristic concept of this product model.

Lemma 2. *Let $k \in \mathbb{N}$.*

1. $X^k(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D)) \in cs_{\mathcal{O}}(C, D)$.
2. *Let $E \in cs_{\mathcal{O}}(C, D)$ with $rd(E) \leq k$.*
   It holds that $X^k(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D)) \sqsubseteq_{\mathcal{O}} E$.

In the proof we only need to refer to the canonical models and its properties given in Lemma 1, thus the proof is a straightforward variant of the one given in [14].

In the following we use $X^k(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ as a representation of the k-lcs$_{\mathcal{O}}(C, D)$. It is implied by Lemma 2 that the set of k-characteristic concepts of the product model $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ for all k is the set of lcs-candidates for the lcs$_{\mathcal{O}}(C, D)$, which can be stated as follows.

Corollary 1. *The $lcs_{\mathcal{O}}(C, D)$ exists iff there exists a $k \in \mathbb{N}$ such that for all $\ell \in \mathbb{N}$: k-$lcs_{\mathcal{O}}(C, D) \sqsubseteq_{\mathcal{O}} \ell$-$lcs_{\mathcal{O}}(C, D)$.*

Obviously, this condition does not yield a decision procedure for the problem whether the k-lcs$_{\mathcal{O}}(C, D)$ is the lcs, since subsumption cannot be checked for infinitely many ℓ in finite time.

Next, we address step 2 and show a condition on the common subsumers that decides whether a common subsumer is least or not. The main idea is that the product model $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ captures all commonalities of the input concepts C and D. Thus we need to compare the canonical models of the common subsumers and the product model by using simulation-equivalence $\simeq$.

Lemma 3. *Let E be a concept. $E \equiv_{\mathcal{O}} lcs_{\mathcal{O}}(C, D)$ iff $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D)) \simeq (\mathcal{I}_{E,\mathcal{O}}, d_E)$.*

Proof sketch. For any $F \in \text{cs}_\mathcal{O}(C,D)$ it holds by Lemma 1, Claim 3 that $(\mathcal{I}_{F,\mathcal{O}}, d_F)$ is simulated by $(\mathcal{I}_{C,\mathcal{O}}, d_C)$ and by $(\mathcal{I}_{D,\mathcal{O}}, d_D)$ and therefore also by $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$.

Assume that $(\mathcal{I}_{E,\mathcal{O}}, d_E)$ is simulation-equivalent to the product model. We need to show that $E \equiv_\mathcal{O} \text{lcs}_\mathcal{O}(C,D)$. By transitivity of $\lesssim$ it is implied that $(\mathcal{I}_{F,\mathcal{O}}, d_F) \lesssim (\mathcal{I}_{E,\mathcal{O}}, d_E)$ and $E \sqsubseteq_\mathcal{O} F$ by Lemma 1. Therefore $E \equiv_\mathcal{O} \text{lcs}_\mathcal{O}(C,D)$.

For the other direction assume $E \equiv_\mathcal{O} \text{lcs}_\mathcal{O}(C,D)$. It has to be shown that $(\mathcal{I}_{E,\mathcal{O}}, d_E)$ simulates the product model. Let $\mathcal{J}_{(d_C,d_D)}$ be the tree unraveling of the product model. Since E is more specific than the k-characteristic concepts of the product model for all k (by Corollary 1), $(\mathcal{I}_{E,\mathcal{O}}, d_E)$ simulates the subtree $\mathcal{J}^k_{(d_C,d_D)}$ of $\mathcal{J}_{(d_C,d_D)}$ limited to elements up to depth k, for all k. For each k we consider the maximal simulation from $\mathcal{J}^k_{(d_C,d_D)}$ to $(\mathcal{I}_{E,\mathcal{O}}, d_E)$. Note that $((d_C, d_D), d_E)$ is contained in any of these simulations. Let σ be an element of $\Delta^{\mathcal{J}_{(d_C,d_D)}}$ at an arbitrary depth ℓ. We show how to determine the elements of $\Delta^{\mathcal{I}_{E,\mathcal{O}}}$, that simulate this fixed element σ. Let $\mathcal{S}_n(\sigma)$ be the maximal set of elements from $\Delta^{\mathcal{I}_{E,\mathcal{O}}}$ that simulate σ in each of the trees $\mathcal{J}^n_{(d_C,d_D)}$ with $n \geq \ell$. We can observe that the infinite sequence $(\mathcal{S}_{\ell+i}(\sigma))_{i=0,1,2,\ldots}$ is decreasing (w.r.t. $\supseteq$). Therefore at a certain depth we reach a fixpoint set. This fixpoint set exists for any σ. It can be shown that the union of all these fixpoint sets yields a simulation from the product model to $(\mathcal{I}_{E,\mathcal{O}}, d_E)$.

By the use of Lemma 3 it can be verified whether a given common subsumer is the least one or not, which we illustrate by an example.

Example 1. Consider the following TBox

$$\begin{aligned}\mathcal{T}_1 = \{ & C \sqsubseteq E \sqcap \exists r.C, \\ & D \sqsubseteq E \sqcap \exists r.D, \\ & E \sqsubseteq \exists s.E\}\end{aligned}$$

with $\mathcal{O}_1 = (\mathcal{T}_1, \emptyset)$ and now the following extended ontology

$$\mathcal{O}_2 = (\mathcal{T}_1, \{s \circ s \sqsubseteq r\})$$

In Figure 1 we can see that

$$E \sqcap \exists s.E \in \text{cs}_{\mathcal{O}_1}(C,D),$$

but this concept is not the lcs of C and D, because its canonical model cannot simulate the product model $(\mathcal{I}_{C,\mathcal{O}_1} \times \mathcal{I}_{D,\mathcal{O}_1}, (d_C, d_D))$. The concept E, however, is the lcs of C and D w.r.t. $\mathcal{O}_2$. We have $(\mathcal{I}_{C,\mathcal{O}_2} \times \mathcal{I}_{D,\mathcal{O}_2}, (d_C, d_D)) \lesssim (\mathcal{I}_{E,\mathcal{O}_2}, d_E)$ since (d_C, d_D) and (d_E, d_E) are simulated by d_E.

The characterization of the existence of the lcs given in Corollary 1 can be reformulated using Lemma 3.

Corollary 2. *The $\text{lcs}_\mathcal{O}(C,D)$ exists iff there exists a k such that the canonical model of $X^k(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ w.r.t. $\mathcal{O}$ simulates $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$.*

This corollary still doesn't yield a decision procedure for the existence problem, since the depth k is still unrestricted. Such a restriction will be developed in the next section.

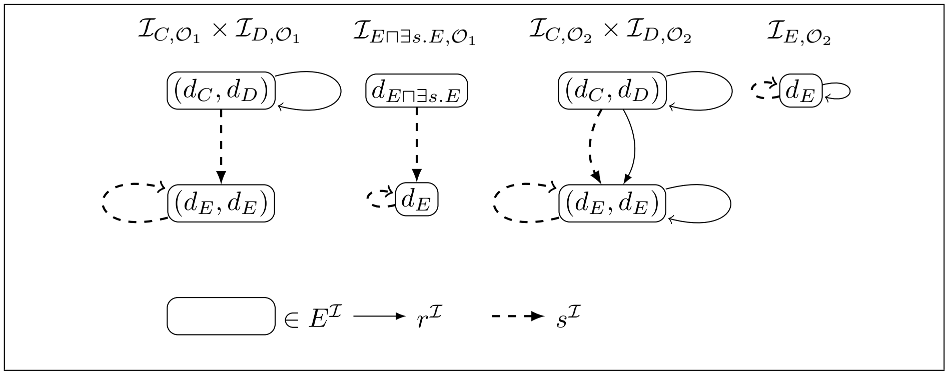

Fig. 1: Product of canonical models of $\mathcal{O}_1$ and $\mathcal{O}_2$

3.2 A Polynomial Upper Bound on the Role-depth of the lcs

In this section we show that, if the lcs exists, its role-depth is bounded by the size of the product model. First, consider again the ontology $\mathcal{O}_2$ from Example 1, where $E \sqsubseteq_{\mathcal{O}_2} \exists r.E$ holds, which results in an r-loop in the product model through the element (d_E, d_E) with the same name appearing in both components. Furthermore, the cycles in the product model involving the roles r and s are exactly those captured by the canonical model $\mathcal{I}_{E,\mathcal{O}_2}$. Therefore $E \equiv_{\mathcal{O}_2} \text{lcs}_{\mathcal{O}_2}(C, D)$. On this observation we build our general method.

We call elements $(d_F, d_{F'}) \in \Delta^{\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}}$ *synchronous* if $F = F'$ and *asynchronous* otherwise. The structure of $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ can now be simplified by considering only synchronous successors of synchronous elements.

In order to find a number k, such that the product model is simulated by the canonical model of $K = X^k(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$, we first represent the model $(\mathcal{I}_{K,\mathcal{O}}, d_K)$ as a subtree of the tree unraveling of the product model $\mathcal{J}_{(d_C,d_D)}$ with root (d_C, d_D). We construct this representation by extending the subtree $\mathcal{J}^k_{(d_C,d_D)}$ by new tree models at depth k. We need to ensure that the resulting interpretation, denoted by $\widehat{\mathcal{J}}^k_{(d_C,d_D)}$, is a model of $\mathcal{O}$, that is simulation-equivalent to $(\mathcal{I}_{K,\mathcal{O}}, d_K)$. The elements $\sigma \in \Delta^{\mathcal{J}^k_{(d_C,d_D)}}$ with $|\sigma| = k$ that we extend and the corresponding trees we append to them are selected as follows: Let M be a conjunction of concept names and $\exists s.F \in \mathsf{sub}(\mathcal{O})$. If $\sigma \in M^{\mathcal{J}^k_{(d_C,d_D)}}$ and $M \sqsubseteq_{\mathcal{O}} \exists r.F$, then we append the tree unraveling of the canonical model $\mathcal{I}_{\exists r.F,\mathcal{O}}$. Note that the role names s and r can be different due to the presence of RIAs in $\mathcal{O}$. Furthermore, we consider elements that have a tail that is a synchronous element. If $\mathsf{tail}(\sigma) = (d_F, d_F)$, then F is called *tail concept* of σ. To select the elements with a synchronous tail, that we extend by the canonical model of their tail concept, we use embeddings of $\mathcal{J}^k_{(d_C,d_D)}$ into $(\mathcal{I}_{K,\mathcal{O}}, d_K)$. Let $\mathcal{H} = \{Z_1, ..., Z_n\}$ be the set of all functional simulations Z_i from $\mathcal{J}^k_{(d_C,d_D)}$ to $(\mathcal{I}_{K,\mathcal{O}}, d_K)$ with $Z_i((d_C, d_D)) = d_K$. We say that σ with tail concept F is *matched* by Z_i if $Z_i(\sigma) \in F^{\mathcal{I}_{K,\mathcal{O}}}$. The set of elements $\sigma \in \Delta^{\mathcal{J}^k_{(d_C,d_D)}}$ with $|\sigma| = k$, that are matched by a functional simulation Z_i is called *matching set*, denoted by $\mathcal{M}(Z_i)$. Now consider the set $\mathcal{M}(\mathcal{H}) := \{\mathcal{M}(Z_1), ..., \mathcal{M}(Z_n)\}$. If σ is contained in *all* maximal matching sets from $\mathcal{M}(\mathcal{H})$, then we extend σ by the tree unraveling of the canonical model of its tail concept w.r.t. $\mathcal{O}$. Intuitively, this condition corresponds to a minimization of changes which is required since the canonical model we want to obtain is minimal w.r.t. $\lesssim$. We can show that the resulting interpretation $\widehat{\mathcal{J}}^k_{(d_C,d_D)}$ has the desired properties.

Lemma 4. *Let $K = X^k(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$. $\widehat{\mathcal{J}}^k_{(d_C,d_D)}$ is a model of $\mathcal{O}$ and $\widehat{\mathcal{J}}^k_{(d_C,d_D)} \simeq (\mathcal{I}_{K,\mathcal{O}}, d_K)$.*

Having this representation of the canonical model of the k-lcs$_\mathcal{O}(C, D)$ at hand, we first show a sufficient condition for the existence of the lcs.

Corollary 3. *If all cycles in $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$, that are reachable from (d_C, d_D) consist of synchronous elements, then the lcs$_\mathcal{O}(C, D)$ exists.*

Proof sketch. There exists an $\ell \in \mathbb{N}$ such that all paths in the tree unraveling $\mathcal{J}_{(d_C,d_D)}$ of $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ starting in (d_C, d_D) have a maximal asynchronous prefix up to length ℓ, i.e., if there exists an element at depth $\geq \ell + 1$, then it is a synchronous element. Consider the number

$$m := \mathsf{max}(\{rd(F) \mid F \in \mathsf{sub}(\mathcal{O}) \cup \{C, D\}\}).$$

We unravel $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ up to depth $\ell + m + 1$ such that we get $\mathcal{J}^{\ell+m+1}_{(d_C,d_D)}$. Now it is ensured that the corresponding model $\widehat{\mathcal{J}}^{\ell+m+1}_{(d_C,d_D)}$ contains all paths with a maximal asynchronous prefix up to length ℓ. It is implied that $\widehat{\mathcal{J}}^{\ell+m+1}_{(d_C,d_D)} = \mathcal{J}_{(d_C,d_D)}$. From Lemma 4 and from Corollary 2 it follows that $X^{\ell+m+1}(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ is the lcs.

As seen in Example 1 for $\mathcal{O}_2$, this is not a necessary condition for the existence of the lcs. Since although the lcs of C and D exists w.r.t. $\mathcal{O}_2$, the product model has also a cycle involving the asynchronous element (d_C, d_D).

Another consequence of Lemma 4 is, that if the product model $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ has only asynchronous cycles reachable from (d_C, d_D), then the lcs$_\mathcal{O}(C, D)$ does not exist. Since in this case $\mathcal{J}_{(d_C,d_D)}$ is infinite but $\widehat{\mathcal{J}}^k_{(d_C,d_D)}$ is finite for all $k \in \mathbb{N}$, a simulation from $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ to $\widehat{\mathcal{J}}^k_{(d_C,d_D)}$ never exists for all k.

The interesting case is where there are both asynchronous and synchronous cycles reachable from (d_C, d_D) in the product model. In this case we choose a k that is large enough and then check whether the corresponding canonical model of $X^k(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ w.r.t. $\mathcal{O}$ simulates the product model.

We show in the next lemma that the role-depth of the lcs$_\mathcal{O}(C, D)$, if it exists, can be bounded by a polynomial, that is quadratic in the size of the product model.

Lemma 5. *Let $m := \mathsf{max}(\{rd(F) \mid F \in \mathsf{sub}(\mathcal{O}) \cup \{C, D\}\})$ and $n := |\Delta^{\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}}|$. If lcs$_\mathcal{O}(C, D)$ exists then $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D)) \lesssim \widehat{\mathcal{J}}^{n^2+m+1}_{(d_C,d_D)}$.*

$$\begin{tikzcd}[row sep=small]
p = & d_0 \arrow[r, "r_1"] \arrow[d, "\mathcal{S}"] & d_1 \arrow[r, "r_2"] \arrow[d, "\mathcal{S}"] & d_2 \arrow[r, "r_3"] \arrow[d, "\mathcal{S}"] & d_3 \arrow[r, "r_4"] \arrow[d, "\mathcal{S}"] & \cdots \\
p_\ell = & \sigma_0 \arrow[r, "r_1"] & \sigma_1 \arrow[r, "r_2"] & \sigma_2 \arrow[r, "r_3"] & \sigma_3 \arrow[r, "r_4"] & \cdots
\end{tikzcd}$$

Fig. 2: simulation chain of p and p_ℓ

Proof sketch. Assume $\mathrm{lcs}_{\mathcal{O}}(C, D)$ exists. From Corollary 2 and Lemma 4 it follows that there exists a number ℓ such that

$$(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D)) \lesssim \widehat{\mathcal{J}}^{\ell}_{(d_C,d_D)}. \tag{1}$$

Every path in $\widehat{\mathcal{J}}^{\ell}_{(d_C,d_D)}$ has a maximal asynchronous prefix of length $\leq \ell$. From depth $\ell + 1$ on there are only synchronous elements in the tree $\widehat{\mathcal{J}}^{\ell}_{(d_C,d_D)}$. From (1) it follows that every path p in $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$ starting in (d_C, d_D), is simulated by a corresponding path p_ℓ in $\widehat{\mathcal{J}}^{\ell}_{(d_C,d_D)}$ also starting in (d_C, d_D). The *simulation chain* of p and p_ℓ is depicted in Figure 2. The idea of a simulation chain is to use the simulating path p_ℓ to construct a simulating path in $\widehat{\mathcal{J}}^{\ell}_{(d_C,d_D)}$ (also starting in (d_C, d_D)) with a maximal asynchronous prefix of length $\leq n^2$, where n^2 is the number of pairs of elements from $\Delta^{\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}}$. Intuitively, if p_ℓ has a maximal asynchronous prefix that is longer than n^2, then there must be pairs in the simulation chain that occur more than once. These are used to construct a simulating path with a shorter maximal asynchronous prefix step-wise. After a finite number of steps the result is a simulating path, such that all pairs consisting of asynchronous elements in the corresponding simulation chain are pairwise distinct. Therefore we need only asynchronous elements from $\widehat{\mathcal{J}}^{\ell}_{(d_C,d_D)}$ up to depth n^2 to simulate the product model. Then we add $m + 1$ to n^2 to ensure that $\widehat{\mathcal{J}}^{n^2+m+1}_{(d_C,d_D)}$ contains *all* paths from $\mathcal{J}_{(d_C,d_D)}$ starting in (d_C, d_D), that have a maximal asynchronous prefix of length $\leq n^2$. As argued above $\widehat{\mathcal{J}}^{n^2+m+1}_{(d_C,d_D)}$ simulates $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$.

Using Lemma 3 and Lemma 5 we can now show the main result of this paper.

Theorem 1. *Let C, D be concepts and $\mathcal{O}$ an $\mathcal{EL}^+$-ontology. It is decidable in polynomial time whether the $lcs_{\mathcal{O}}(C, D)$ exists. If the $lcs_{\mathcal{O}}(C, D)$ exists then it can be computed in polynomial time.*

Proof. First we compute the bound k as given in Lemma 5 and then the k-characteristic concept K of $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D))$. The canonical model of K can be build according to Definition 2 in polynomial time [5]. Next we check whether $(\mathcal{I}_{C,\mathcal{O}} \times \mathcal{I}_{D,\mathcal{O}}, (d_C, d_D)) \lesssim (\mathcal{I}_{K,\mathcal{O}}, d_K)$ holds, which can be done in polynomial time. If yes, K is the lcs by Lemma 3 and if no, the lcs doesn't exist by Lemma 5.

The results from this section can be easily generalized to the lcs of an arbitrary set of concepts $M = \{C_1, ..., C_m\}$ w.r.t. an ontology $\mathcal{O}$. But in this case the size of the lcs is already exponential w.r.t. an empty TBox [2]. In this general case we have to take the product model

$$(\mathcal{I}_{C_1,\mathcal{O}} \times \cdots \times \mathcal{I}_{C_m,\mathcal{O}}, (d_{C_1}, \cdots, d_{C_m})),$$

whose size is exponential in the size of M and $\mathcal{O}$, as input for the methods introduced in this section. Then the same steps as for the binary version can be applied.

4 Conclusions and Future Work

In this paper we have studied the conditions for the existence of the lcs, if computed w.r.t. $\mathcal{EL}^+$-ontologies. In this setting the lcs doesn't need to exist. We showed that the existence problem of the lcs of two concepts is decidable in polynomial time. Furthermore, we showed that the role-depth of the lcs can be bounded by a polynomial. This upper bound k can be used to compute the lcs, if it exists. Otherwise the computed concept can still serve as an approximation [8].

Future work on the practical side includes to improve the described procedure in order to obtain a practical algorithm such that an appropriate implementation can be integrated into existing tools [7] for computing generalizations. On the theoretical side, we would like extend the results towards the computation of most specific concepts of individuals w.r.t. to ontologies consisting of general TBoxes and cyclic ABoxes. Furthermore, we will also consider these generalization inferences w.r.t. ontologies formulated in more expressive Horn-DLs than $\mathcal{EL}^+$.

References

1. Baader, F., Calvanese, D., McGuinness, D., Nardi, D., Patel-Schneider, P. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications (2003)
2. Baader, F., Küsters, R., Molitor, R.: Computing least common subsumers in description logics with existential restrictions. pp. 96–101. Stockholm, Sweden (1999)
3. Baader, F.: Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In: Gottlob, G., Walsh, T. (eds.) Proceedings of the 18th International Joint Conference on Artificial Intelligence. pp. 319–324. Morgan Kaufman (2003)
4. Baader, F.: A graph-theoretic generalization of the least common subsumer and the most specific concept in the description logic $\mathcal{EL}$. In: Hromkovic, J., Nagl, M. (eds.) Proceedings of the 30th International Workshop on Graph-Theoretic Concepts in Computer Science (WG 2004). Lecture Notes in Computer Science, vol. 3353, pp. 177–188. Springer-Verlag (2004)
5. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05. Morgan-Kaufmann Publishers (2005)

6. Baader, F., Sertkaya, B., Turhan, A.-Y.: Computing the least common subsumer w.r.t. a background terminology. Journal of Applied Logic 5(3), 392–420 (2007)
7. Ecke, A., Turhan, A.-Y.: Optimizations for the role-depth bounded least common subsumer in $\mathcal{EL}^+$. In: Klinov, P., Horridge, M. (eds.) Proceedings of OWL: Experiences and Directions Workshop. CEUR Workshop Proceedings, vol. 849. CEUR-WS.org (2012)
8. Ecke, A., Turhan, A.-Y.: Role-depth bounded least common subsumers for $\mathcal{EL}^+$ and $\mathcal{ELI}$. In: Kazakhov, Y., Wolter, F. (eds.) Proc. of Description Logics Workshop. CEUR, vol. 846 (2012)
9. Lutz, C., Piro, R., Wolter, F.: Enriching $\mathcal{EL}$-concepts with greatest fixpoints. In: Proceedings of the 19th European Conference on Artificial Intelligence (ECAI10). IOS Press (2010)
10. Lutz, C., Wolter, F.: Deciding inseparability and conservative extensions in the description logic $\mathcal{EL}$. Journal of Symbolic Computation 45(2), 194–228 (2010)
11. Peñaloza, R., Turhan, A.-Y.: A practical approach for computing generalization inferences in $\mathcal{EL}$. In: Grobelnik, M., Simperl, E. (eds.) Proceedings of the 8th European Semantic Web Conference (ESWC'11). Lecture Notes in Computer Science, Springer-Verlag (2011)
12. Pokrywczynski, D., Walther, D.: Deciding the logical difference problem for $\mathcal{EL}$ with role inclusions. In: Proc. of the International Workshop on Ontologies: Reasoning and Modularity (WORM'08), Tenerife, Spain. CEUR-WS Vol-348 (2008)
13. Zarrieß, B., Turhan, A.-Y.: Most specific generalizations w.r.t. general $\mathcal{EL}$-TBoxes. In: Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI'13). AAAI Press, Beijing, China (2013), to appear.
14. Zarrieß, B., Turhan, A.-Y.: Most specific generalizations w.r.t. general $\mathcal{EL}$-TBoxes. LTCS-Report 13-06, Chair of Automata Theory, Institute of Theoretical Computer Science, Technische Universität Dresden, Dresden, Germany (2013), see `http://lat.inf.tu-dresden.de/research/reports.html`.