TECHNISCHE
UNIVERSITÄT
DRESDEN

Dresden University of Technology
Institute for Theoretical Computer Science
Chair for Automata Theory

LTCS–Report

Least common subsumers, most specific concepts, and role-value-maps in a description logic with existential restrictions and terminological cycles

Franz Baader

LTCS-Report 02-07

Lehrstuhl für Automatentheorie
Institut für Theoretische Informatik
TU Dresden
http://lat.inf.tu-dresden.de

Hans-Grundig-Str. 25
01062 Dresden
Germany

Least common subsumers, most specific concepts, and role-value-maps in a description logic with existential restrictions and terminological cycles

Franz Baader*
Theoretical Computer Science
Dresden University of Technology
D-01062 Dresden, Germany
e-mail: baader@inf.tu-dresden.de

Abstract

In a previous report we have investigates subsumption in the presence of terminological cycles for the description logic $\mathcal{EL}$, which allows conjunctions, existential restrictions, and the top concept, and have shown that the subsumption problem remains polynomial for all three types of semantics usually considered for cyclic definitions in description logics. This result depends on a characterization of subsumption through the existence of certain simulation relations on the graph associated with a terminology.

In the present report we will use this characterization to show how the most specific concept and the least common subsumer can be computed in $\mathcal{EL}$ with cyclic definitions. In addition, we show that subsumption in $\mathcal{EL}$ (with or without cyclic definitions) remains polynomial even if one adds a certain restricted form of global role-value-maps to $\mathcal{EL}$. In particular, this kind of role-value-maps can express transitivity of roles.

1 Introduction

Computing the most specific concept of an individual and the least common subsumer of concepts can be used in the bottom-up construction of description logic (DL) knowledge bases. Instead of defining the relevant concepts of an application domain from scratch, this methodology allows the user to give typical examples of individuals belonging to the concept to be defined. These individuals are then

*Partially supported by the DFG under grant BA 1122/4-1.

generalized to a concept by first computing the most specific concept of each individual (i.e., the least concept description in the available description language that has this individual as an instance), and then computing the least common subsumer of these concepts (i.e., the least concept description in the available description language that subsumes all these concepts). The knowledge engineer can then use the computed concept as a starting point for the concept definition.

The least common subsumer (lcs) in DLs with existential restrictions was investigated in [3]. In particular, it was shown there that the lcs in the small DL $\mathcal{EL}$ (which allows conjunctions, existential restrictions, and the top concept) always exists, and that the binary lcs can be computed in polynomial time. Unfortunately, the most specific concept (msc) of a given ABox individual need not exist in languages allowing for existential restrictions or number restrictions. As a possible solution to this problem, Küsters and Molitor [9] show how the most specific concept can be approximated in $\mathcal{EL}$ and some of its extensions. Here, we follow an alternative approach: we extend the language by cyclic terminologies with greatest fixpoint semantics, and show that the msc always exists in this setting. For the DL $\mathcal{ALN}$ (which allows conjunctions, value restrictions, and number restrictions) it was shown in [2] that the most specific concept always exists if one adds cyclic concept definitions with gfp-semantics. One reason for Küsters and Molitor to choose an approximation approach rather than an exact characterization of the most specific concept using cyclic definitions was that the impact of cyclic definitions in description logics with existential restrictions was largely unexplored.

The report [1] was a first step toward overcoming this deficit. It considers cyclic terminologies in $\mathcal{EL}$ w.r.t. the three types of semantics (greatest fixpoint, least fixpoint, and descriptive semantics) introduced by Nebel [12], and shows that the subsumption problem can be decided in polynomial time in all three cases. This is in stark contrast to the case of DLs with value restrictions. Even for the small DL $\mathcal{FL}_0$ (which allows conjunctions and value restrictions only), adding cyclic terminologies increases the complexity of the subsumption problem from polynomial (for concept descriptions) to PSPACE. The main tool in the investigation of cyclic definitions in $\mathcal{EL}$ is a characterization of subsumption through the existence of so-called simulation relations, which can be computed in polynomial time [7]. The results in [1] also show that cyclic definitions with least fixpoint semantics are not interesting in $\mathcal{EL}$. For this reason, we will here concentrate on greatest fixpoint and descriptive semantics.

The characterization of subsumption in $\mathcal{EL}$ w.r.t. gfp-semantics through the existence of certain simulation relations on the graph associated with the terminology can be used to characterize the lcs via the product of this graph with itself (Section 4.1). This shows that, w.r.t. gfp semantics, the lcs always exists, and the binary lcs can be computed in polynomial time. (The n-ary lcs may grow exponentially even in $\mathcal{EL}$ without cyclic terminologies [3].) For cyclic terminologies in

$\mathcal{EL}$ with descriptive semantics, the lcs need not exist (Section 4.2). We introduce possible candidates P_k ($k \geq 0$) for the lcs, and show that the lcs exists iff one of these candidates is the lcs. In addition, we give a sufficient condition for the lcs to exist, and show that, under this condition, it can be computed in polynomial time.

The characterization of subsumption w.r.t. gfp-semantics can be extended to the instance problem in $\mathcal{EL}$. This allows us to show that the msc in $\mathcal{EL}$ with cyclic terminologies interpreted with gfp semantics always exists, and can be computed in polynomial time (Section 5).

In Section 6, we extend the results of [1] in another direction. In many applications (e.g., in medicine [15] and in process engineering [13]), one uses roles that are not just arbitrary binary relations, but should satisfy certain relationships. A prominent example are transitive roles r, which satisfy $r \circ r \sqsubseteq r$, i.e., the composition of r with itself is a subrelation of r. In Section 6 we consider more general constraints of the form $r_1 \circ r_2 \sqsubseteq r_3$, which say that the composition of r_1 with r_2 is a subrelation of r_3. Obviously, this is a special form of role-value-maps [14], which are global in the sense that they must hold for every individual in the interpretation domain. The right-identity rule in [15] is a special case where r_1 is identical with r_3. As an example, consider the roles $\mathsf{location}$, which assigns objects with their location, and $\mathsf{contained}$, which relates each spacial region with those regions containing it. Then it makes sense to assert the condition $\mathsf{location} \circ \mathsf{contained} \sqsubseteq \mathsf{location}$. We will show that adding global role-value-maps of the form $r_1 \circ r_2 \sqsubseteq r_3$ to $\mathcal{EL}$ with cyclic terminologies (interpreted with gfp or descriptive semantics) leaves the subsumption problem polynomial. In particular, this shows that subsumption of $\mathcal{EL}$-concept descriptions (with or without acyclic terminologies) remains polynomial when adding these global role-value-maps.

In the next section, we introduce $\mathcal{EL}$ with cyclic terminologies as well as the lcs and the msc. Then we recall the important definitions and results from [1]. Section 4 formulates and proves the new results for the lcs, and Section 5 does the same for the msc. Finally, Section 6 is devoted to showing the results for global role-value-maps mentioned above.

2 Cyclic terminologies, least common subsumers, and most specific concepts

Concept descriptions are inductively defined with the help of a set of *constructors*, starting with a set N_C of *concept names* and a set N_R of *role names*. The constructors determine the expressive power of the DL. In this report, we restrict the attention to the DL $\mathcal{EL}$, whose concept descriptions are formed using the constructors top-concept ($\top$), conjunction ($C \sqcap D$), and existential restriction

name of constructor	Syntax	Semantics
concept name $A \in N_C$	A	$A^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}}$
role name $r \in N_R$	r	$r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$
top-concept	$\top$	$\Delta^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y : (x, y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
concept definition	$A \equiv D$	$A^{\mathcal{I}} = D^{\mathcal{I}}$
individual name $a \in N_I$	a	$a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$
concept assertion	$A(a)$	$a^{\mathcal{I}} \in \mathcal{A}^{\mathcal{I}}$
role assertion	$r(a, b)$	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$

Table 1: Syntax and semantics of $\mathcal{EL}$-concept descriptions, TBox definitions, and ABox assertions.

($\exists r.C$). The semantics of $\mathcal{EL}$-concept descriptions is defined in terms of an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. The domain $\Delta^{\mathcal{I}}$ of $\mathcal{I}$ is a non-empty set of individuals and the interpretation function $\cdot^{\mathcal{I}}$ maps each concept name $A \in N_C$ to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ and each role $r \in N_R$ to a binary relation $r^{\mathcal{I}}$ on $\Delta^{\mathcal{I}}$. The extension of $\cdot^{\mathcal{I}}$ to arbitrary concept descriptions is inductively defined, as shown in the third column of Table 1.

A *terminology* (or *TBox* for short) is a finite set of concept definitions of the form $A \equiv D$, where A is a concept name and D a concept description. In addition, we require that TBoxes do not contain *multiple definitions*, i.e., there cannot be two distinct concept descriptions D_1 and D_2 such that both $A \equiv D_1$ and $A \equiv D_2$ belongs to the TBox. Concept names occurring on the left-hand side of a definition are called *defined concepts*. All other concept names occurring in the TBox are called *primitive concepts*. Note that we allow for cyclic dependencies between the defined concepts, i.e., the definition of A may refer (directly or indirectly) to A itself. An interpretation $\mathcal{I}$ is a model of the TBox $\mathcal{T}$ iff it satisfies all its concept definitions, i.e., $A^{\mathcal{I}} = D^{\mathcal{I}}$ for all definitions $A \equiv D$ in $\mathcal{T}$.

An *ABox* is a finite set of assertions of the form $A(a)$ and $r(a, b)$, where A is a concept name, r is a role name, and a, b are individual names from a set N_I. Interpretations of ABoxes must additionally map each individual name $a \in N_I$ to an element $a^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$. An interpretation $\mathcal{I}$ is a model of the ABox $\mathcal{A}$ iff it satisfies all its assertions, i.e., $a^{\mathcal{I}} \in A^{\mathcal{I}}$ for all concept assertions $A(a)$ in $\mathcal{A}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ for all role assertions $r(a, b)$ in $\mathcal{A}$. The interpretation $\mathcal{I}$ is a model of the ABox $\mathcal{A}$ together with the TBox $\mathcal{T}$ iff it is a model of both $\mathcal{T}$ and $\mathcal{A}$.

The semantics of (possibly cyclic) $\mathcal{EL}$-TBoxes we have defined above is called *descriptive semantic* by Nebel [12]. For some applications, it is more appropriate to interpret cyclic concept definitions with the help of an appropriate fixpoint semantics.

Example 1 To illustrate this, let us recall an example from [1]:

$$\mathsf{Inode} \equiv \mathsf{Node} \sqcap \exists \mathsf{edge}.\mathsf{Inode}.$$

Here the intended interpretations are graphs where we have nodes (elements of the concept Node) and edges (represented by the role edge), and we want to define the concept Inode of all nodes lying on an infinite (possibly cyclic) path of the graph. In order to capture this intuition, the above definition must be interpreted with greatest fixpoint semantics.

Before we can define greatest fixpoint semantics (gfp-semantics), we must introduce some notation. Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox containing the roles N_{role}, the primitive concepts N_{prim}, and the defined concepts $N_{def} = \{A_1, \ldots, A_k\}$. A *primitive interpretations* $\mathcal{J}$ for $\mathcal{T}$ is given by a domain $\Delta^{\mathcal{J}}$, an interpretation of the roles $r \in N_{role}$ by binary relations $r^{\mathcal{J}}$ on $\Delta^{\mathcal{J}}$, and an interpretation of the primitive concepts $P \in N_{prim}$ by subsets $P^{\mathcal{J}}$ of $\Delta^{\mathcal{J}}$. Obviously, a primitive interpretation differs from an interpretation in that it does not interpret the defined concepts in N_{def}. We say that the interpretation $\mathcal{I}$ is *based on* the primitive interpretation $\mathcal{J}$ iff it has the same domain as $\mathcal{J}$ and coincides with $\mathcal{J}$ on N_{role} and N_{prim}. For a fixed primitive interpretation $\mathcal{J}$, the interpretations $\mathcal{I}$ based on it are uniquely determined by the tuple $(A_1^{\mathcal{I}}, \ldots, A_k^{\mathcal{I}})$ of the interpretations of the defined concepts in N_{def}. We define

$$\mathit{Int}(\mathcal{J}) := \{\mathcal{I} \mid \mathcal{I} \text{ is an interpretation based on } \mathcal{J}\}.$$

Interpretations based on $\mathcal{J}$ can be compared by the following ordering, which realizes a pairwise inclusion test between the respective interpretations of the defined concepts: if $\mathcal{I}_1, \mathcal{I}_2 \in \mathit{Int}(\mathcal{J})$, then

$$\mathcal{I}_1 \preceq_{\mathcal{J}} \mathcal{I}_2 \text{ iff } A_i^{\mathcal{I}_1} \subseteq A_i^{\mathcal{I}_2} \text{ for all } i, 1 \leq i \leq k.$$

It is easy to see that $\preceq_{\mathcal{J}}$ is a *complete lattice* on $\mathit{Int}(\mathcal{J})$, i.e., every subset of $\mathit{Int}(\mathcal{J})$ has a least upper bound (lub) and a greatest lower bound (glb). Thus, *Tarski's fixpoint theorem* [18, 10] applies to all monotonic functions from $\mathit{Int}(\mathcal{J})$ to $\mathit{Int}(\mathcal{J})$. This theorem states the following: if $O\colon \mathit{Int}(\mathcal{J}) \to \mathit{Int}(\mathcal{J})$ is a function such that $\mathcal{I}_1 \preceq_{\mathcal{J}} \mathcal{I}_2$ implies $O(\mathcal{I}_1) \preceq_{\mathcal{J}} O(\mathcal{I}_2)$ (monotonicity), then O has a *fixpoint*, i.e., there is an $\mathcal{I}$ in $\mathit{Int}(\mathcal{J})$ such that $O(\mathcal{I}) = \mathcal{I}$. In particular, it has a greatest fixpoint, i.e., a fixpoint larger w.r.t. $\preceq_{\mathcal{J}}$ than all other fixpoints.

Definition 2 The TBox $\mathcal{T} := \{A_1 \equiv D_1, \ldots, A_k \equiv D_k\}$ induces the following function $O_{\mathcal{T},\mathcal{J}}$ on $\mathit{Int}(\mathcal{J})$: $O_{\mathcal{T},\mathcal{J}}(\mathcal{I}_1) = \mathcal{I}_2$ iff $A_i^{\mathcal{I}_2} = D_i^{\mathcal{I}_1}$ holds for all $i, 1 \leq i \leq k$.

It is easy to see that, for a given $\mathcal{EL}$-TBox $\mathcal{T}$ and a primitive $\mathcal{J}$, the function $O_{\mathcal{T},\mathcal{J}}$ is indeed monotonic. Consequently, $O_{\mathcal{T},\mathcal{J}}$ has a greatest fixpoint. It is an immediate consequence of the definition of $O_{\mathcal{T},\mathcal{J}}$ that an interpretation $\mathcal{I}$ based on the primitive interpretation $\mathcal{J}$ is a fixpoint of $O_{\mathcal{T},\mathcal{J}}$ iff $\mathcal{I}$ is a model of $\mathcal{T}$. This shows that any primitive interpretation $\mathcal{J}$ can be extended to a model of $\mathcal{T}$. In particular, there is always a greatest model of $\mathcal{T}$ extending $\mathcal{J}$.

Definition 3 Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox. The model $\mathcal{I}$ of $\mathcal{T}$ is called *gfp-model* of $\mathcal{T}$ iff there is a primitive interpretation $\mathcal{J}$ such that $\mathcal{I} \in Int(\mathcal{J})$ is the greatest fixpoint of $O_{\mathcal{T},\mathcal{J}}$. *Greatest fixpoint semantics* considers only gfp-models as admissible models.

We are now ready to define the subsumption and the instance problem w.r.t. the two different types of semantics introduced above.

Definition 4 Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and $\mathcal{A}$ an $\mathcal{EL}$-ABox, let A, B be defined concepts occurring in $\mathcal{T}$, and a an individual name occurring in $\mathcal{A}$. Then,

- A is subsumed by B w.r.t. descriptive semantics ($A \sqsubseteq_{\mathcal{T}} B$) iff $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ holds for all models $\mathcal{I}$ of $\mathcal{T}$.
- a is an instance of A w.r.t. descriptive semantics ($\mathcal{A} \models_{\mathcal{T}} A(a)$) iff $a^{\mathcal{I}} \in A^{\mathcal{I}}$ holds for all models $\mathcal{I}$ of $\mathcal{T}$ together with $\mathcal{A}$.
- A is subsumed by B w.r.t. gfp-semantics ($A \sqsubseteq_{gfp,\mathcal{T}} B$) iff $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ holds for all gfp-models $\mathcal{I}$ of $\mathcal{T}$.
- a is an instance of A w.r.t. gfp-semantics ($\mathcal{A} \models_{gfp,\mathcal{T}} A(a)$) iff $a^{\mathcal{I}} \in A^{\mathcal{I}}$ holds for all models $\mathcal{I}$ of $\mathcal{A}$ that are gfp-models of $\mathcal{T}$.

On the level of concept descriptions, the least common subsumer of two concept descriptions C, D is the least concept description E that subsumes both C and D. An extensions of this definition to the level of (possibly cyclic) TBoxes is not completely trivial. In fact, assume that A_1, A_2 are concepts defined in the TBox $\mathcal{T}$. It should be obvious that taking as the lcs of A_1, A_2 the least defined concept B in $\mathcal{T}$ such that $A_1 \sqsubseteq_{\mathcal{T}} B$ and $A_2 \sqsubseteq_{\mathcal{T}} B$ is too weak since the lcs would then strongly depend on what other defined concepts are already present in $\mathcal{T}$. However, a second approach (which might look like the obvious generalization of the definition of the lcs in the case of concept descriptions) is also not quite satisfactory (at least if we consider gfp-semantics). We could say that the lcs of A, B is the least concept description C (possibly using defined concepts of $\mathcal{T}$) such that $A_1 \sqsubseteq_{\mathcal{T}} C$ and $A_2 \sqsubseteq_{\mathcal{T}} C$ (respectively, $A_1 \sqsubseteq_{gfp,\mathcal{T}} C$ and $A_2 \sqsubseteq_{gfp,\mathcal{T}} C$). The problem is that this definition does not allow us to use the expressive power of cyclic definitions (with gfp-semantics) when constructing the lcs. For example, consider the TBox $\mathcal{T}$ consisting of the following concept definitions:

$$\begin{aligned} \mathsf{Bluelnode} &\equiv \mathsf{Blue} \sqcap \mathsf{Node} \sqcap \exists \mathsf{edge}.\mathsf{Bluelnode}, \\ \mathsf{Redlnode} &\equiv \mathsf{Red} \sqcap \mathsf{Node} \sqcap \exists \mathsf{edge}.\mathsf{Redlnode}. \end{aligned}$$

The intended interpretation is similar to the one in Example 1, with the only difference that now nodes may have colors, and we are interested in blue (red) nodes lying on an infinite path consisting of blue (red) nodes. Intuitively, the lcs of $\mathsf{BlueInode}$ and $\mathsf{RedInode}$ describes nodes lying on an infinite path (without any restriction on their color), i.e., the concept Inode from Example 1 should be a definition of this lcs. However, this cannot be expressed by a simple concept description. It requires a new cyclic definition.

Consequently, to obtain the lcs we must allow the original TBox to be extended by new definitions. We say that the TBox $\mathcal{T}_2$ is a *conservative extension* of the TBox $\mathcal{T}_1$ iff $\mathcal{T}_1 \subseteq \mathcal{T}_2$ and $\mathcal{T}_1$ and $\mathcal{T}_2$ have the same primitive concepts and roles. Thus, $\mathcal{T}_2$ may contain new definitions $A \equiv D$, but then D does not introduce new primitive concepts and roles (i.e., all of them already occur in $\mathcal{T}_1$), and A is a new concept name (i.e., A does not occur in $\mathcal{T}_1$). The name "conservative extension" is justified by the fact that the new definitions in $\mathcal{T}_2$ do not influence the subsumption relationships between defined concepts in $\mathcal{T}_1$.

Lemma 5 *Let $\mathcal{T}_1, \mathcal{T}_2$ be $\mathcal{EL}$-TBoxes such that $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$, and let A, B be defined concepts in $\mathcal{T}_1$ (and thus also in $\mathcal{T}_2$). Then $A \sqsubseteq_{\mathcal{T}_1} B$ iff $A \sqsubseteq_{\mathcal{T}_2} B$. The same holds for subsumption w.r.t. gfp-semantics.*

Proof. (1) Let us first consider descriptive semantics. The implication from left to right ($\Rightarrow$) is trivial since $\mathcal{T}_1 \subseteq \mathcal{T}_2$ (monotonicity of first-order logic).

For the other direction ($\Leftarrow$), one should note that $\mathcal{T} := \mathcal{T}_2 \setminus \mathcal{T}_1$ can be viewed as a TBox whose primitive concepts are the defined and primitive concepts of $\mathcal{T}_1$, and whose roles are the roles of $\mathcal{T}_1$. Now, assume that $A \not\sqsubseteq_{\mathcal{T}_1} B$, and let $\mathcal{I}$ be a model of $\mathcal{T}_1$ such that $A^{\mathcal{I}} \not\subseteq B^{\mathcal{I}}$. The model $\mathcal{I}$ of $\mathcal{T}_1$ can be viewed as a primitive interpretation of $\mathcal{T}$, which can be extended to a gfp-model $\widehat{\mathcal{I}}$ of $\mathcal{T}$. Obviously, $\widehat{\mathcal{I}}$ is also a model of $\mathcal{T}_2$, and since it coincides with $\mathcal{I}$ on the primitive and defined concepts in $\mathcal{T}_1$, it also satisfies $A^{\widehat{\mathcal{I}}} = A^{\mathcal{I}} \not\subseteq B^{\mathcal{I}} = B^{\widehat{\mathcal{I}}}$.

(2) Now, let us consider gfp-semantics. The implication from right to left ($\Leftarrow$) can be proved similar to ($\Leftarrow$) of part (1) of the proof (where now we start with a gfp-model $\mathcal{I}$ of $\mathcal{T}_1$). What remains to be shown is that $\widehat{\mathcal{I}}$ is a *gfp*-model of $\mathcal{T}_2$. Thus, assume that there is a larger model $\mathcal{I}'$ of $\mathcal{T}_2$ based on the same primitive interpretation. The difference between $\widehat{\mathcal{I}}$ and $\mathcal{I}'$ cannot occur on one of the defined concepts of $\mathcal{T}_1$ since this would contradict our assumption that $\mathcal{I}$ is a gfp-model of $\mathcal{T}_1$. Consequently, the restriction of $\mathcal{I}'$ to the defined concepts in $\mathcal{T}_1$ coincides with $\mathcal{I}$. But then a difference between $\mathcal{I}'$ and $\widehat{\mathcal{I}}$ in one of the concepts newly defined in $\mathcal{T}_2$ contradicts the fact that $\widehat{\mathcal{I}}$ is a gfp-model of $\mathcal{T}$ (see part (1) of the proof).

The implication from left to right ($\Rightarrow$) immediately follows if we can show that the restriction $\mathcal{I}'$ of a gfp-model $\mathcal{I}$ of $\mathcal{T}_2$ to the defined concepts of $\mathcal{T}_1$ is a gfp-model of $\mathcal{T}_1$. Obviously, $\mathcal{I}'$ is a model of $\mathcal{T}_1$ (for being a restriction of a model of

$\mathcal{T}_2$). Now, assume that it is not a gfp-model of $\mathcal{T}_1$. Thus, there is a larger model $\mathcal{I}''$ of $\mathcal{T}_1$ that coincides with $\mathcal{I}'$ on the primitive concepts and roles. As in ($\Rightarrow$) of part (2) of the proof, we can show that $\mathcal{I}''$ can be extended to a gfp-model of $\mathcal{T}_2$. However, this gfp-model is based on the same primitive interpretation as $\mathcal{I}$, and thus must be identical to $\mathcal{I}$, which contradicts our assumption that $\mathcal{I}''$ is larger than $\mathcal{I}'$. $\square$

Definition 6 Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox containing the defined concepts A, B, and let $\mathcal{T}_2$ be a conservative extension of $\mathcal{T}_1$ containing the new defined concept E. Then E in $\mathcal{T}_2$ is a *least common subsumer* of A, B in $\mathcal{T}_1$ *w.r.t. descriptive semantics* (lcs) iff the following two conditions are satisfied:

1. $A \sqsubseteq_{\mathcal{T}_2} E$ and $B \sqsubseteq_{\mathcal{T}_2} E$.
2. If $\mathcal{T}_3$ is a conservative extension of $\mathcal{T}_2$ and F a defined concept in $\mathcal{T}_3$ such that $A \sqsubseteq_{\mathcal{T}_3} F$ and $B \sqsubseteq_{\mathcal{T}_3} F$, then $E \sqsubseteq_{\mathcal{T}_3} F$.

Least common subsumers w.r.t. gfp-semantics (gfp-lcs) are defined analogously, by replacing $\sqsubseteq_{\mathcal{T}_i}$ by $\sqsubseteq_{gfp,\mathcal{T}_i}$.

In the case of concept descriptions, the lcs is unique up to equivalence, i.e., if E_1 and E_2 are both least common subsumers of the descriptions C, D, then $E_1 \equiv E_2$ (i.e., $E_1 \sqsubseteq E_2$ and $E_2 \sqsubseteq E_1$). In the presence of (possibly acyclic) TBoxes, this uniqueness property also holds (though its formulation is more complicated).

Proposition 7 *Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox containing the defined concepts A, B. Assume that $\mathcal{T}_2$ and $\mathcal{T}_2'$ are conservative extensions of $\mathcal{T}_1$ such that*

- *the defined concept E in $\mathcal{T}_2$ is an lcs of A, B in $\mathcal{T}_1$;*
- *the defined concept E' in $\mathcal{T}_2'$ is an lcs of A, B in $\mathcal{T}_1$;*
- *the sets of newly defined concepts in respectively $\mathcal{T}_2$ and $\mathcal{T}_2'$ are disjoint.*

Where $\mathcal{T}_3 := \mathcal{T}_2 \cup \mathcal{T}_2'$, we have $E \equiv_{\mathcal{T}_3} E'$ (i.e., $E \sqsubseteq_{\mathcal{T}_3} E'$ and $E' \sqsubseteq_{\mathcal{T}_3} E$).

The corresponding statement holds for the gfp-lcs.

Proof. Since the sets of newly defined concepts in respectively $\mathcal{T}_2$ and $\mathcal{T}_2'$ are disjoint, $\mathcal{T}_3 := \mathcal{T}_2 \cup \mathcal{T}_2'$ is a conservative extension of both $\mathcal{T}_2$ and $\mathcal{T}_2'$. Consequently $A \sqsubseteq_{\mathcal{T}_2} E$ and $B \sqsubseteq_{\mathcal{T}_2} E$ imply $A \sqsubseteq_{\mathcal{T}_3} E$ and $B \sqsubseteq_{\mathcal{T}_3} E$, and $A \sqsubseteq_{\mathcal{T}_2'} E'$ and $B \sqsubseteq_{\mathcal{T}_2'} E'$ imply $A \sqsubseteq_{\mathcal{T}_3} E$ and $B \sqsubseteq_{\mathcal{T}_3} E$. Since E in $\mathcal{T}_2$ is an lcs of A, B, this implies that $E \sqsubseteq_{\mathcal{T}_3} E'$. Analogously, since E' in $\mathcal{T}_2'$ is an lcs of A, B, this implies that $E' \sqsubseteq_{\mathcal{T}_3} E$.

The same argument goes through for the gfp-lcs. □

The notion "most specific concept" can be extended in a similar way from concept descriptions to concepts defined in a TBox.

Definition 8 Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox and $\mathcal{A}$ an $\mathcal{EL}$-ABox containing the individual name a, and let $\mathcal{T}_2$ be a conservative extension of $\mathcal{T}_1$ containing the defined concept E. Then E in $\mathcal{T}_2$ is a *most specific concept* of a in $\mathcal{A}$ and $\mathcal{T}_1$ *w.r.t. descriptive semantics* (msc) iff the following two conditions are satisfied:

1. $\mathcal{A} \models_{\mathcal{T}_2} E(a)$.
2. If $\mathcal{T}_3$ is a conservative extension of $\mathcal{T}_2$ and F a defined concept in $\mathcal{T}_3$ such that $\mathcal{A} \models_{\mathcal{T}_3} F(a)$, then $E \sqsubseteq_{\mathcal{T}_3} F$.

Most specific concepts w.r.t. gfp-semantics (gfp-msc) are defined analogously.

Uniqueness up to equivalence of the most specific concept can be shown like uniqueness of the least common subsumer.

Proposition 9 *Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox and $\mathcal{A}$ an $\mathcal{EL}$-ABox containing the individual name a. Assume that $\mathcal{T}_2$ and $\mathcal{T}_2'$ are conservative extensions of $\mathcal{T}_1$ such that*

- *the defined concept E in $\mathcal{T}_2$ is an msc of a in $\mathcal{A}$ and $\mathcal{T}_1$;*
- *the defined concept E' in $\mathcal{T}_2'$ is an msc of a in $\mathcal{A}$ and $\mathcal{T}_1$;*
- *the sets of newly defined concepts in respectively $\mathcal{T}_2$ and $\mathcal{T}_2'$ are disjoint.*

Where $\mathcal{T}_3 := \mathcal{T}_2 \cup \mathcal{T}_2'$, we have $E \equiv_{\mathcal{T}_3} E'$.

The corresponding statement holds for the gfp-msc.

3 Characterizing subsumption in $\mathcal{EL}$ with cyclic definitions

In this section, we recall the characterizations of subsumption w.r.t. descriptive semantics and gfp-semantics developed in [1]. To this purpose, we must represent TBoxes by description graphs, and introduce the notion of a simulation on description graphs.

3.1 Description graphs and simulations

It was shown in [1] that $\mathcal{EL}$-TBoxes as well as primitive interpretations can be represented as description graphs. Before we can translate $\mathcal{EL}$-TBoxes into description graphs, we must normalize the TBoxes. In the following, let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, N_{def} the defined concepts of $\mathcal{T}$, N_{prim} the primitive concepts of $\mathcal{T}$, and N_{role} the roles of $\mathcal{T}$.

We say that the $\mathcal{EL}$-TBox $\mathcal{T}$ is *normalized* iff $A \equiv D \in \mathcal{T}$ implies that D is of the form

$$P_1 \sqcap \ldots \sqcap P_m \sqcap \exists r_1.B_1 \sqcap \ldots \sqcap \exists r_\ell.B_\ell,$$

for $m, \ell \geq 0$, $P_1, \ldots, P_m \in N_{prim}$, $r_1, \ldots, r_\ell \in N_{role}$, and $B_1, \ldots, B_\ell \in N_{def}$. If $m = \ell = 0$, then $D = \top$.

As shown in [1], one can (without loss of generality) restrict the attention to normalized TBox. In the following, we thus assume that all TBoxes are normalized. Normalized $\mathcal{EL}$-TBoxes can be viewed as graphs whose nodes are the defined concepts, which are labeled by sets of primitive concepts, and whose edges are given by the existential restrictions. For the rest of this section, we fix a normalized $\mathcal{EL}$-TBox $\mathcal{T}$ with primitive concepts N_{prim}, defined concepts N_{def}, and roles N_{role}.

Definition 10 An *$\mathcal{EL}$-description graph* is a graph $\mathcal{G} = (V, E, L)$ where

- V is a set of nodes;
- $E \subseteq V \times N_{role} \times V$ is a set of edges labeled by role names;
- $L\colon V \to 2^{N_{prim}}$ is a function that labels nodes with sets of primitive concepts.

The TBox $\mathcal{T}$ can be translated into the following $\mathcal{EL}$-description graph $\mathcal{G}_\mathcal{T} = (N_{def}, E_\mathcal{T}, L_\mathcal{T})$:

- the nodes of $\mathcal{G}_\mathcal{T}$ are the defined concepts of $\mathcal{T}$;
- if A is a defined concept and

  $$A \equiv P_1 \sqcap \ldots \sqcap P_m \sqcap \exists r_1.B_1 \sqcap \ldots \sqcap \exists r_\ell.B_\ell$$

  its definition in $\mathcal{T}$, then

  – $L_\mathcal{T}(A) = \{P_1, \ldots, P_m\}$, and
  – A is the source of the edges $(A, r_1, B_1), \ldots, (A, r_\ell, B_\ell) \in E_\mathcal{T}$.

Any primitive interpretation $\mathcal{J} = (\Delta^\mathcal{J}, \cdot^\mathcal{J})$ can be translated into the following $\mathcal{EL}$-description graph $\mathcal{G}_\mathcal{J} = (\Delta^\mathcal{J}, E_\mathcal{J}, L_\mathcal{J})$:

- the nodes of $\mathcal{G}_\mathcal{J}$ are the elements of $\Delta^\mathcal{J}$;
- $E_\mathcal{J} := \{(x, r, y) \mid (x, y) \in r^\mathcal{J}\}$;
- $L_\mathcal{J}(x) = \{P \in N_{prim} \mid x \in P^\mathcal{J}\}$ for all $x \in \Delta^\mathcal{J}$.

Simulations are binary relations between nodes of two $\mathcal{EL}$-description graphs that respect labels and edges in the sense defined below.

Definition 11 Let $\mathcal{G}_i = (V_i, E_i, L_i)$ $(i = 1, 2)$ be two $\mathcal{EL}$-description graphs. The binary relation $Z \subseteq V_1 \times V_2$ is a *simulation* from $\mathcal{G}_1$ to $\mathcal{G}_2$ iff

(S1) $(v_1, v_2) \in Z$ implies $L_1(v_1) \subseteq L_2(v_2)$; and

(S2) if $(v_1, v_2) \in Z$ and $(v_1, r, v_1') \in E_1$, then there exists a node $v_2' \in V_2$ such that $(v_1', v_2') \in Z$ and $(v_2, r, v_2') \in E_2$.

We write $Z\colon \mathcal{G}_1 \overset{\rightharpoonup}{\sim} \mathcal{G}_2$ to express that Z is a simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$.

It is easy to see that the set of all simulations from $\mathcal{G}_1$ to $\mathcal{G}_2$ is closed under arbitrary unions. Consequently, there always exists a greatest simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$. If $\mathcal{G}_1, \mathcal{G}_2$ are finite, then this greatest simulation can be computed in polynomial time [7]. As an easy consequence of this fact, the following proposition is proved in [1].

Proposition 12 *Let $\mathcal{G}_1, \mathcal{G}_2$ be two finite $\mathcal{EL}$-description graphs, v_1 a node of $\mathcal{G}_1$ and v_2 a node of $\mathcal{G}_2$. Then we can be decide in polynomial time whether there is a simulation $Z\colon \mathcal{G}_1 \overset{\rightharpoonup}{\sim} \mathcal{G}_2$ such that $(v_1, v_2) \in Z$.*

3.2 Subsumption w.r.t. gfp-semantics

Subsumption w.r.t. gfp-semantics corresponds to the existence of a simulation relation such that the subsumee simulates the subsumer:

Theorem 13 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and A, B defined concepts in $\mathcal{T}$. Then the following are equivalent:*

1. *$A \sqsubseteq_{gfp,\mathcal{T}} B$.*
2. *There is a simulation $Z\colon \mathcal{G}_\mathcal{T} \overset{\rightharpoonup}{\sim} \mathcal{G}_\mathcal{T}$ such that $(B, A) \in Z$.*

The theorem together with Proposition 12 shows that subsumption w.r.t. gfp-semantics in $\mathcal{EL}$ is tractable.

Corollary 14 *Subsumption w.r.t. gfp-semantics in $\mathcal{EL}$ can be decided in polynomial time.*

This result is quite surprising since, for the DL $\mathcal{FL}_0$ (which allows for conjunction and value restrictions only), subsumption w.r.t. gfp-semantics is already PSPACE-complete.

The proof of the above theorem given in [1] depends on a characterization of when an individual of a gfp-model belongs to a defined concept in this model.

Proposition 15 *Let $\mathcal{J}$ be a primitive interpretation and $\mathcal{I}$ the gfp-model of $\mathcal{T}$ based on $\mathcal{J}$. Then the following are equivalent for any $A \in N_{def}$ and $x \in \Delta^{\mathcal{J}}$:*

1. *$x \in A^{\mathcal{I}}$.*

2. *There is a simulation Z: $\mathcal{G}_{\mathcal{T}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{J}}$ such that $(A, x) \in Z$.*

This proposition will become relevant later on when we extend the characterization of subsumption to a characterization of the instance problem.

3.3 Subsumption w.r.t. descriptive semantics

Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and $\mathcal{G}_{\mathcal{T}}$ the corresponding $\mathcal{EL}$-description graph. Since every gfp-model of $\mathcal{T}$ is a model of $\mathcal{T}$, $A \sqsubseteq_{\mathcal{T}} B$ implies $A \sqsubseteq_{gfp,\mathcal{T}} B$. Consequently, $A \sqsubseteq_{\mathcal{T}} B$ implies that there is a simulation Z: $\mathcal{G}_{\mathcal{T}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}}$ with $(B, A) \in Z$. However, the simulation Z must satisfy some additional properties for the implication in the other direction to hold. To define these properties, we must introduce some notation.

Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{G}_{\mathcal{T}}$ the corresponding $\mathcal{EL}$-description graph, and Z: $\mathcal{G}_{\mathcal{T}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}}$ a simulation.

Definition 16 The path p_1: $B = B_0 \overset{r_1}{\rightarrow} B_1 \overset{r_2}{\rightarrow} B_2 \overset{r_3}{\rightarrow} B_3 \overset{r_4}{\rightarrow} \cdots$ in $\mathcal{G}_{\mathcal{T}}$ is Z-*simulated* by the path p_2: $A = A_0 \overset{r_1}{\rightarrow} A_1 \overset{r_2}{\rightarrow} A_2 \overset{r_3}{\rightarrow} A_3 \overset{r_4}{\rightarrow} \cdots$ in $\mathcal{G}_{\mathcal{T}}$ iff $(B_i, A_i) \in Z$ for all $i \geq 0$. In this case we say that the pair (p_1, p_2) is a (B, A)*-simulation chain w.r.t. Z*. (see Figure 1).

If $(B, A) \in Z$, then (S2) of Definition 11 implies that, for every infinite path p_1 starting with $B_0 := B$, there is an infinite path p_2 starting with $A_0 := A$ such that p_1 is Z-simulated by p_2. In the following we construct such a simulating path step by step. The main point is, however, that the decision which concept A_n to take in step n should depend only on the partial (B, A)-simulation chain already constructed, and *not* on the parts of the path p_1 not yet considered.

$$\begin{array}{ccccccccccc} B & = & B_0 & \stackrel{r_1}{\to} & B_1 & \stackrel{r_2}{\to} & B_2 & \stackrel{r_3}{\to} & B_3 & \stackrel{r_4}{\to} & \cdots \\ & & Z\downarrow & & Z\downarrow & & Z\downarrow & & Z\downarrow & & \\ A & = & A_0 & \stackrel{r_1}{\to} & A_1 & \stackrel{r_2}{\to} & A_2 & \stackrel{r_3}{\to} & A_3 & \stackrel{r_4}{\to} & \cdots \end{array}$$

Figure 1: A (B, A)-simulation chain.

$$\begin{array}{ccccccccccc} B & = & B_0 & \stackrel{r_1}{\to} & B_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & B_{n-1} & \stackrel{r_n}{\to} & B_n \\ & & Z\downarrow & & Z\downarrow & & & & Z\downarrow & & \\ A & = & A_0 & \stackrel{r_1}{\to} & A_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & A_{n-1} & & \end{array}$$

Figure 2: A partial (B, A)-simulation chain.

Definition 17 A *partial* (B, A)*-simulation chain* is of the form depicted in Figure 2. A *selection function* S for A, B and Z assigns to each partial (B, A)-simulation chain of this form a defined concept A_n such that (A_{n-1}, r_n, A_n) is an edge in $\mathcal{G}_\mathcal{T}$ and $(B_n, A_n) \in Z$.

Given a path $B = B_0 \stackrel{r_1}{\to} B_1 \stackrel{r_2}{\to} B_2 \stackrel{r_3}{\to} B_3 \stackrel{r_4}{\to} \cdots$ and a defined concept A such that $(B, A) \in Z$, one can use a selection function S for A, B and Z to construct a Z-simulating path. In this case we say that the resulting (B, A)-simulation chain is S*-selected.*

Definition 18 Let A, B be defined concepts in $\mathcal{T}$, and $Z\colon \mathcal{G}_\mathcal{T} \overset{\rightarrow}{\sim} \mathcal{G}_\mathcal{T}$ a simulation with $(B, A) \in Z$. Then Z is called (B, A)*-synchronized* iff there exists a selection function S for A, B and Z such that the following holds: for every infinite S-selected (B, A)-simulation chain of the form depicted in Figure 1 there exists an $i \geq 0$ such that $A_i = B_i$.

We are now ready to state the characterization of subsumption w.r.t. descriptive semantics proved in [1].

Theorem 19 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, and A, B defined concepts in $\mathcal{T}$. Then the following are equivalent:*

1. $A \sqsubseteq_\mathcal{T} B$.
2. *There is a (B, A)-synchronized simulation $Z\colon \mathcal{G}_\mathcal{T} \overset{\rightarrow}{\sim} \mathcal{G}_\mathcal{T}$ such that $(B, A) \in Z$.*

In [1] it is also shown that, for a given $\mathcal{EL}$-TBox $\mathcal{T}$ and defined concepts A, B in $\mathcal{T}$, the existence of a (B, A)-synchronized simulation $Z\colon \mathcal{G}_\mathcal{T} \overset{\rightarrow}{\sim} \mathcal{G}_\mathcal{T}$ with $(B, A) \in Z$ can be decided in polynomial time.

Corollary 20 *Subsumption w.r.t. descriptive semantics in $\mathcal{EL}$ can be decided in polynomial time.*

4 Computing the lcs

We will first show how the characterization of subsumption w.r.t. gfp-semantics given in Theorem 13 can be used to characterize the gfp-lcs. Deriving a characterization of the lcs (w.r.t. descriptive semantics) from Theorem 19 turns out to be more involved.

4.1 Computing the gfp-lcs

Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox, let $\mathcal{G}_{\mathcal{T}_1} = (N_{def}, E_{\mathcal{T}_1}, L_{\mathcal{T}_1})$ be the corresponding description graph, and let A, B be defined concepts in $\mathcal{T}_1$ (i.e., elements of N_{def}). In principle, the lcs of A, B in $\mathcal{T}_1$ is defined in a TBox whose description graph is the product of $\mathcal{G}_{\mathcal{T}_1}$ with itself.

Definition 21 Let $\mathcal{G}_1 = (V_1, E_1, L_1)$ and $\mathcal{G}_2 = (V_2, E_2, L_2)$ be two description graphs. Their *product* is the description graph $\mathcal{G}_1 \times \mathcal{G}_2 := (V, E, L)$ where

- $V = V_1 \times V_2$;
- $E := \{((v_1, v_2), r, (v_1', v_2')) \mid (v_1, r, v_1') \in E_1 \wedge (v_2, r, v_2') \in E_2\}$;
- $L(v_1, v_2) := L_1(v_1) \cap L_2(v_2)$.

The description graph $\mathcal{G}_{\mathcal{T}_1} \times \mathcal{G}_{\mathcal{T}_1}$ yields a TBox $\mathcal{T}$ such that $\mathcal{G}_{\mathcal{T}} = \mathcal{G}_{\mathcal{T}_1} \times \mathcal{G}_{\mathcal{T}_1}$. Now, $\mathcal{T}_2 := \mathcal{T}_1 \cup \mathcal{T}$ is a conservative extension of $\mathcal{T}_1$. In fact, $\mathcal{G}_{\mathcal{T}_1} \times \mathcal{G}_{\mathcal{T}_1}$ (and thus $\mathcal{T}$) is based on the same primitive concepts and roles as $\mathcal{G}_{\mathcal{T}_1}$, and the set of defined concepts in $\mathcal{T}$ is $N_{def} \times N_{def}$, which is disjoint from N_{def}. Let $\mathcal{G}_2 = (V_2, E_2, L_2)$ be the $\mathcal{EL}$-description graph corresponding to $\mathcal{T}_2$. Note that $\mathcal{G}_2$ is the disjoint union of $\mathcal{G}_{\mathcal{T}} = \mathcal{G}_{\mathcal{T}_1} \times \mathcal{G}_{\mathcal{T}_1}$ and $\mathcal{G}_{\mathcal{T}_1}$. Let $\mathcal{G}_{\mathcal{T}} = (V, E, L)$ and $\mathcal{G}_{\mathcal{T}_1} = (V_1, E_1, L_1)$.

Lemma 22 *(A, B) in $\mathcal{T}_2$ is the gfp-lcs of A and B in $\mathcal{T}_1$.*

Proof. (1) First, we show that $A \sqsubseteq_{gfp,\mathcal{T}_2} (A, B)$. (Note that $B \sqsubseteq_{gfp,\mathcal{T}_2} (A, B)$ can be shown analogously.) According to Theorem 13 it is sufficient to show that there exists a simulation relation Z: $\mathcal{G}_{\mathcal{T}_2} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}_2}$ such that $((A, B), A) \in Z$. We define Z as the projection of elements of $N_{def} \times N_{def}$ to the first component, i.e.,

$$Z := \{((u, v), u) \mid (u, v) \in N_{def} \times N_{def}\}.$$

Note that the nodes $(u, v) \in N_{def} \times N_{def}$ are exactly the defined concepts of $\mathcal{T}$.

Obviously, $((A, B), A) \in Z$ by definition of Z. It remains to be shown that Z is a simulation relation:

(S1) By the definition of the product of $\mathcal{EL}$-description graphs, $L_2(u, v) = L(u, v) = L_1(u) \cap L_1(v) \subseteq L_1(u) = L_2(u)$.

(S2) Consider $((u, v), u) \in Z$ and assume that $((u, v), r, w) \in E_2$ for some node $w \in V_2$. Since $\mathcal{G}_2$ is the disjoint union of $\mathcal{G}_{\mathcal{T}} = \mathcal{G}_{\mathcal{T}_1} \times \mathcal{G}_{\mathcal{T}_1}$ and $\mathcal{G}_{\mathcal{T}_1}$, and (u, v) is a node of $\mathcal{G}_{\mathcal{T}}$, w must also be a node of $\mathcal{G}_{\mathcal{T}}$, i.e., w is of the form (u', v') and the edge $((u, v), r, (u', v')) \in E_2$ is an edge in $\mathcal{G}_{\mathcal{T}}$. Thus, $((u, v), r, (u', v')) \in E$, and the definition of the product of $\mathcal{EL}$-description graphs implies that $(u, r, u') \in E_1 \subseteq E_2$. Since $((u', v'), u') \in Z$, this shows that property (S2) in the definition of simulation relations really holds for Z.

(2) Now, assume that $\mathcal{T}_3$ is a conservative extension of $\mathcal{T}_2$ and that F is a defined concept in $\mathcal{T}_3$ such that $A \sqsubseteq_{gfp,\mathcal{T}_3} F$ and $B \sqsubseteq_{gfp,\mathcal{T}_3} F$. Where $\mathcal{G}_{\mathcal{T}_3} = (V_3, E_3, L_3)$, this implies that there are simulation relations $Y_1\colon \mathcal{G}_{\mathcal{T}_3} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}_3}$ and $Y_2\colon \mathcal{G}_{\mathcal{T}_3} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}_3}$ such that $(F, A) \in Y_1$ and $(F, B) \in Y_2$.

We must show that $(A, B) \sqsubseteq_{gfp,\mathcal{T}_3} F$, i.e., that there is a simulation relation $Y\colon \mathcal{G}_{\mathcal{T}_3} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}_3}$ such that $(F, (A, B)) \in Y$. Basically, Y is defined as the "product" of Y_1 and Y_2. To be more precise,

$$Y := \{(u, (v_1, v_2)) \mid (u, v_1) \in Y_1 \wedge (u, v_2) \in Y_2 \wedge (v_1, v_2) \in V = N_{def} \times N_{def}\}.$$

Since $(F, A) \in Y_1$ and $(F, B) \in Y_2$, and $(A, B) \in V = N_{def} \times N_{def}$, we know that $(F, (A, B)) \in Y$. It remains to be shown that Y is in fact a simulation relation.

(S1) Assume that $(u, (v_1, v_2)) \in Y$, i.e., $(u, v_1) \in Y_1$, $(u, v_1) \in Y_2$, and $(v_1, v_2) \in V$. Since Y_1 and Y_2 are simulation relations, the first two facts imply that $L_3(u) \subseteq L_3(v_1)$ and $L_3(u) \subseteq L_3(v_2)$, and thus $L_3(u) \subseteq L_3(v_1) \cap L_3(v_2)$. Since $(v_1, v_2) \in V$ and $\mathcal{T}_3$ is a conservative extension of $\mathcal{T}_2$, we have for $i = 1, 2$: $L_3(v_i) = L_2(v_i) = L_1(v_i)$. By the definition of the product, this implies $L_3(u) \subseteq L_3(v_1) \cap L_3(v_2) = L_1(v_1) \cap L_1(v_2) = L(v_1, v_2) = L_3(v_1, v_2)$.

(S2) Assume that $(u, (v_1, v_2)) \in Y$ and that $(u, r, u') \in E_3$. By the definition of Y, and since Y_1 and Y_2 are simulation relations, there exist nodes v_1' and v_2' in V_3 such that $(v_1, r, v_1') \in E_3$, $(u', v_1') \in Y_1$, $(v_2, r, v_2') \in E_3$, and $(u', v_2') \in Y_2$. Again by the definition of Y, v_1, v_2 are nodes in $V_1 = N_{def}$. By the definition of $\mathcal{T}_2$, and since $\mathcal{T}_3$ is a conservative extension of $\mathcal{T}_2$, this implies that the edges (v_1, r, v_1') and (v_2, r, v_2') are actually edges in E_1, and thus the definition of the product yields $((v_1, v_2), r, (v_1', v_2')) \in E \subseteq E_3$. In addition, this shows that $(v'_1, v'_2) \in V$, and thus $(u', v'_1) \in Y_1$ and $(u', v'_2) \in Y_2$ imply that $(u', (v'_1, v'_2)) \in Y$. $\square$

Computing the (binary) product of two $\mathcal{EL}$-description graphs can obviously be done in polynomial time, and thus the gfp-lcs can be computed in polynomial time.

Theorem 23 *Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox, and let A, B be defined concepts in $\mathcal{T}_1$. Then the gfp-lcs of A, B in $\mathcal{T}_1$ always exists, and it can be computed in polynomial time.*

4.2 The lcs w.r.t. descriptive semantics

First, we will show that, w.r.t. descriptive semantics, the lcs of two concepts defined in an $\mathcal{EL}$-TBox need not exist. Subsequently, we will introduce possible "candidates" P_k $(k \geq 0)$ for the lcs, and show that the lcs exists iff one of these candidates is the lcs. Finally, we will give a sufficient condition for the existence of the lcs.

4.2.1 The lcs need not exist

Theorem 24 *Let $\mathcal{T}_1 := \{A \equiv \exists r.A,\ B \equiv \exists r.B\}$. Then, A, B in $\mathcal{T}_1$ do not have an lcs.*

Proof. Assume to the contrary that $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ and that the defined concept E in $\mathcal{T}_2$ is an lcs of A, B in $\mathcal{T}_1$. Let $\mathcal{G}_2 = (V_2, E_2, L_2)$ be the description graph induced by $\mathcal{T}_2$.

First, we show that there cannot be an infinite path in $\mathcal{G}_2$ starting with E. In fact, assume that

$$E = E_0 \xrightarrow{r_1} E_1 \xrightarrow{r_2} E_2 \xrightarrow{r_3} \cdots$$

is such an infinite path. Since $A \sqsubseteq_{\mathcal{T}_1} E$, there is an (E, A)-synchronized simulation Z_1: $\mathcal{G}_2 \overset{\rightarrow}{\sim} \mathcal{G}_2$ such that $(E, A) \in Z_1$. Consequently, the corresponding selection function S_1 can be used to turn the above infinite chain issuing from E into an (E, A)-simulation chain. Since the only edge with source A is the edge (A, r, A), this simulation chain is actually of the form

$$\begin{array}{ccccccccccc} E & = & E_0 & \xrightarrow{r} & E_1 & \xrightarrow{r} & E_2 & \xrightarrow{r} & E_3 & \xrightarrow{r} & \cdots \\ & & Z_1\downarrow & & Z_1\downarrow & & Z_1\downarrow & & Z_1\downarrow & & \\ & & A & \xrightarrow{r} & A & \xrightarrow{r} & A & \xrightarrow{r} & A & \xrightarrow{r} & \cdots \end{array}$$

Since Z_1 is (E, A)-synchronized with selection function S_1, this implies that there is an index j_1 such that $E_{j_1} = A$, and thus $E_i = A$ for all $i \geq j_1$.

Analogously, we can show that there is an index j_2 such that $E_{j_2} = B$, and thus $E_i = B$ for all $i \geq j_2$. Since $A \neq B$, this is a contradiction. Thus, we know that there is a positive integer n_0 such that every path in $\mathcal{G}_2$ starting with E has length $\leq n_0$.

Second, we define conservative extensions $\mathcal{T}'_n$ $(n \geq 1)$ of $\mathcal{T}_2$ such that the defined concept F_n in $\mathcal{T}'_n$ is a common subsumer of A, B:

$$\mathcal{T}'_n := \mathcal{T}_2 \cup \{F_n \equiv \exists r.F_{n-1}, ..., F_1 \equiv \exists r.F_0, F_0 \equiv \top\}.$$

It is easy to see that $A \sqsubseteq_{\mathcal{T}'_n} F_n$ and $B \sqsubseteq_{\mathcal{T}'_n} F_n$.

Third, we claim that, for $n > n_0$, $E \not\sqsubseteq_{\mathcal{T}'_n} F_n$. In fact, the path

$$F_n \xrightarrow{r} F_{n-1} \xrightarrow{r} F_{n-2} \xrightarrow{r} \cdots \xrightarrow{r} F_0$$

has length n, and thus it cannot be simulated by any path starting with E. This shows that $E \not\sqsubseteq_{\mathcal{T}'_n} F_n$, and thus contradicts our assumption that E in $\mathcal{T}_2$ is the lcs of A, B in $\mathcal{T}_1$. □

4.2.2 Characterizing when the lcs exists

Given an $\mathcal{EL}$-TBox $\mathcal{T}_1$ and defined concepts A, B in $\mathcal{T}_1$, we will defined for each $k \geq 0$ a conservative extension $\mathcal{T}_2^{(k)}$ of $\mathcal{T}_1$ containing a defined concept P_k, and show that A, B have an lcs iff there is a k such that P_k is the lcs of A, B.

To prove this result, we will need a sleight modification of Theorem 19. However, this modified theorem follows easily from the the proof of Theorem 19 given in [1]. Recall that a selection function S for A, B and Z assigns to each partial (B, A)-simulation chain of the form depicted in Figure 2 a defined concept A_n such that (A_{n-1}, r_n, A_n) is an edge in $\mathcal{G}_\mathcal{T}$ and $(B_n, A_n) \in Z$.

Definition 25 We call a selection function S *nice* iff it satisfies the following two conditions:

1. It is *memoryless*, i.e., its result A_n depends only on $B_{n-1}, A_{n-1}, r_n, B_n$, and not on the other parts of the partial (B, A)-simulation chain.
2. If $B_{n-1} = A_{n-1}$, then its result A_n is just B_n.

The simulation relation Z is called *strongly (B, A)-synchronized* iff there exists a *nice* selection function S for A, B and Z such that the following holds: for every infinite S-selected (B, A)-simulation chain of the form depicted in Figure 1 there exists an $i \geq 0$ such that $A_i = B_i$.

Corollary 26 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, and A, B be defined concepts in $\mathcal{T}$. Then the following are equivalent:*

1. $A \sqsubseteq_{\mathcal{T}} B$.
2. *There is a* strongly *(B, A)-synchronized simulation Z: $\mathcal{G}_{\mathcal{T}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}}$ such that $(B, A) \in Z$.*

Proof. $(2 \Rightarrow 1)$ is an immediate consequence of Theorem 19.

$(1 \Rightarrow 2)$ follows from the fact that the simulation relation Y defined in the proof of $(1 \Rightarrow 2)$ of Theorem 19 (see page 27 in [1]) is strongly (B, A)-synchronized. In fact, it is easy to check that the selection function S defined in the proof of Lemma 35 in [1] is indeed nice. □

Strongly (B, A)-synchronized simulations satisfy the following property:

Lemma 27 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox containing at most n defined concepts, let A, B be defined concepts in $\mathcal{T}$, and let Z: $\mathcal{G}_{\mathcal{T}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}}$ be a strongly (B, A)-synchronized simulation relation. Consider an infinite S-selected (B, A)-simulation chain of the form depicted in Figure 1. Then there exists an $m < n^2$ such that $B_m = A_m$.*

Proof. Consider the $n^2 + 1$ tuples $(A_0, B_0), \dots, (A_{n^2}, B_{n^2})$. By definition of n, there are at most n^2 different tuples of this kind, and thus there exist indices $0 \le i < j \le n^2$ such that $(B_i, A_i) = (B_j, A_j)$. Since S is memoryless, the following is also an S-selected simulation chain:

$$\begin{array}{ccccccccccccc}
B = B_0 & \overset{r_1}{\to} & \cdots & \overset{r_i}{\to} & B_j = B_i & \overset{r_{i+1}}{\to} & B_{i+1} & \overset{r_{i+2}}{\to} & \cdots & \overset{r_j}{\to} & B_j = B_i & \overset{r_{i+1}}{\to} & \cdots \\
Z\downarrow & & & & Z\downarrow & & Z\downarrow & & & & Z\downarrow & & \\
A = A_0 & \overset{r_1}{\to} & \cdots & \overset{r_i}{\to} & A_j = A_i & \overset{r_{i+1}}{\to} & A_{i+1} & \overset{r_{i+2}}{\to} & \cdots & \overset{r_j}{\to} & A_j = A_i & \overset{r_{i+1}}{\to} & \cdots
\end{array}$$

Now, the fact that this chain must be synchronized shows that there is indeed an $m < j \le n^2$ such that $B_m = A_m$. □

Obviously, the lemma also holds for finite S-selected (B, A)-simulation chains, provided that they are long enough, i.e., of length at least n^2.

Now, let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox, let $\mathcal{G}_{\mathcal{T}_1} = (N_{def}, E_{\mathcal{T}_1}, L_{\mathcal{T}_1})$ be the corresponding description graph, and let A, B be defined concepts in $\mathcal{T}_1$ (i.e., elements of N_{def}). We consider the product $\mathcal{G} := \mathcal{G}_{\mathcal{T}_1} \times \mathcal{G}_{\mathcal{T}_1}$ of $\mathcal{G}_{\mathcal{T}_1}$ with itself. Let $\mathcal{G} = (V, E, L)$.

The product graph $\mathcal{G}$ as a whole cannot be part of the lcs of A, B since it may contain cycles reachable from (A, B), which would prevent the subsumption relationship between A and (A, B) to hold. Nevertheless, the lcs must "contain" paths in $\mathcal{G}$ starting with (A, B) up to a certain length k. In order to obtain these paths without also getting the cycles in $\mathcal{G}$, we make copies of the nodes in $\mathcal{G}$ on levels between 1 and k. Actually, we will not need nodes of the form (u, u) since they are represented by the nodes u in $\mathcal{G}_{\mathcal{T}_1}$.

To be more precise, we define

$$\mathcal{P}_k := \{(A, B)^0\} \cup \{(u, v)^n \mid u \neq v, (u, v) \in N_{def} \times N_{def} \text{ and } 1 \leq n \leq k\}.$$

For $p = (u, v)^n \in \mathcal{P}$ we call (u, v) the node of p and n the level of p.

The edges of $\mathcal{G}$ induce edges between elements of $\mathcal{P}_k$. To be more precise, we define the set of edges $E_{\mathcal{P}_k}$ as follows: $(p, r, q) \in E_{\mathcal{P}_k}$ iff the following conditions are satisfied:

- $p, q \in \mathcal{P}_k$;
- $p = (u, v)^n$ for some $n, 0 \leq n \leq k$;
- $q = (u', v')^{n+1}$;
- $((u, v), r, (u', v')) \in E$;

Note that the graph $(\mathcal{P}_k, E_{\mathcal{P}_k})$ is a directed acyclic graph. The only element on level 0 is $(A, B)^0$.

The label of an element of $\mathcal{P}_k$ is the label of its node in $\mathcal{G}$, i.e., if $p = (u, v)^n \in \mathcal{P}$, then

$$L_{\mathcal{P}_k}(p) := L(u, v) = L_1(u) \cap L_1(v).$$

We are now ready to define an $\mathcal{EL}$-description graph $\mathcal{G}_2^{(k)}$ whose corresponding TBox $\mathcal{T}_2^{(k)}$ is a conservative extension of $\mathcal{T}_1$, and which contains a defined concept P_k that is a common subsumer of A, B.

Definition 28 For all $k \geq 0$, we define $\mathcal{G}_2^{(k)} := (V_2^{(k)}, E_2^{(k)}, L_2^{(k)})$ where

- $V_2^{(k)} := N_{def} \cup \mathcal{P}_k$;
- $L_2^{(k)} = L_1 \cup L_{\mathcal{P}_k}$, i.e.

$$L_2^{(k)}(v) := \begin{cases} L_1(v) & \text{if } v \in N_{def} \\ L_{\mathcal{P}_k}(v) & \text{if } v \in \mathcal{P}_k \end{cases}$$

- $E_2^{(k)}$ consists of the edges in E_1 and $E_{\mathcal{P}_k}$, extended by some additional edges from $\mathcal{P}_k$ to N_{def}:

$$E_2^{(k)} := E_1 \cup E_{\mathcal{P}} \cup \{(p, r, w) \mid p = (u, v)^n \in \mathcal{P}_k \text{ and } (u, r, w) \in E_1 \text{ and } (v, r, w) \in E_1\}.$$

Let $\mathcal{T}_2^{(k)}$ be the $\mathcal{EL}$-TBox such that $\mathcal{G}_2^{(k)} = \mathcal{G}_{\mathcal{T}_2^{(k)}}$. It is easy to see that $\mathcal{T}_2^{(k)}$ is a conservative extension of $\mathcal{T}_1$.

Lemma 29 *$A \sqsubseteq_{\mathcal{T}_2^{(k)}} (A, B)^0$ and $B \sqsubseteq_{\mathcal{T}_2^{(k)}} (A, B)^0$.*

Proof. We prove $A \sqsubseteq_{\mathcal{T}_2^{(k)}} (A, B)^0$. (The other subsumption relationship can be shown analogously.)

According to Theorem 19, it is enough to show that there is an $((A, B)^0, A)$-synchronized simulation $Z\colon \mathcal{G}_{\mathcal{T}_2^{(k)}} \stackrel{\rightharpoonup}{\sim} \mathcal{G}_{\mathcal{T}_2^{(k)}}$ such that $((A, B)^0, A) \in Z$. We define this simulation relation as follows:

$$\begin{aligned} Z := \{&(p, u) \mid p \in \mathcal{P}_k, u \in N_{def}, \text{ and the node of } p \text{ is of the form } (u, v)\} \cup \\ &\{(u, u) \mid u \in N_{def}\}. \end{aligned}$$

First, note that obviously $((A, B)^0, A) \in Z$.

Second, we show that Z is indeed a simulation relation, i.e., it satisfies (S1) and (S2) of Definition 11.

(S1) First, consider $(p, u) \in Z$ for some $p \in \mathcal{P}_k$. If (u, v) is the node of p, then $L_2^{(k)}(p) = L_{\mathcal{P}_k}(p) = L_1(u) \cap L_1(v) \subseteq L_1(u) = L_2^{(k)}(u)$. The case $(u, u) \in Z$ is trivial.

(S2) For the case $(u, u) \in Z$, this property is trivially satisfied. Now, consider $(p, u) \in Z$ for $p \in \mathcal{P}_k$ and $u \in N_{def}$, and let (u, v) be the node of p.

Case 1: $(p, r, q) \in E_2^{(k)}$ for some $q \in \mathcal{P}_k$.
Consequently,

$$p = (u, v)^n \quad \text{and} \quad q = (u', v')^{n+1}$$

for two distinct nodes $u', v' \in N_{def}$ and some $n, 0 \leq n < k$. The definitions of $E_{\mathcal{P}_k}$ and E imply that $(u, r, u') \in E_1 \subseteq E_2^{(k)}$. In addition, $(q, u') \in Z$ by definition of Z.

Case 2: $(p, r, u') \in E_2^{(k)}$ for some $u' \in N_{def}$.
Recall that (u, v) is the node of p. By definition of $E_2^{(k)}$, $(p, r, u') \in E_2^{(k)}$ implies that $(u, r, u') \in E_1 \subseteq E_2^{(k)}$, and by the definition of Z we have $(u', u') \in Z$.

To sum up, we have shown that Z is a simulation relation such that $((A, B)^0, A) \in Z$. It remains to be shown that Z is $((A, B)^0, A)$-synchronized. Our proof of (S2) yields the desired selection function:

- In the situation $(p, r, q) \in E_2^{(k)}$ and $(p, u) \in Z$, S takes the first component of the node of q.
- In the situation $(p, r, u') \in E_2^{(k)}$ and $(p, u) \in Z$, S takes u'.

- In the situation $(u, r, v) \in E_2^{(k)}$ and $(u, u) \in Z$, S takes v.

Why does S satisfy the synchronization property? Since the directed acyclic graph $(\mathcal{P}_k, E_{\mathcal{P}_k})$ only contains paths of length $\leq k$, any infinite $((A, B)^0, A)$-simulation chain must contain nodes from N_{def} also in the upper component. Restricted to these nodes in the first component, Z is the identity relation. $\square$

What we want show next is that every common subsumer of A, B also subsumes $(A, B)^0$ in $\mathcal{T}_2^{(k)}$ for an appropriate k.

To make this more precise, assume that $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$, and that F is a defined concept in $\mathcal{T}_2$ such that $A \sqsubseteq_{\mathcal{T}_2} F$ and $B \sqsubseteq_{\mathcal{T}_2} F$. Where $\mathcal{G}_{\mathcal{T}_2} = (V_2, E_2, L_2)$, this implies that there is

- an (F, A)-synchronized simulation relation $Y_1 \colon \mathcal{G}_{\mathcal{T}_2} \mathrel{\vec{\sim}} \mathcal{G}_{\mathcal{T}_2}$ with selection function S_1 such that $(F, A) \in Y_1$, and
- an (F, B)-synchronized simulation relation $Y_2 \colon \mathcal{G}_{\mathcal{T}_2} \mathrel{\vec{\sim}} \mathcal{G}_{\mathcal{T}_2}$ with selection function S_2 such that $(F, B) \in Y_2$.

By Corollary 26 we may assume without loss of generality that the selection functions S_1, S_2 are nice. Consequently, if $k = |V_2|^2$, then Lemma 27 shows that the selection functions S_1, S_2 ensure synchronization after less than k steps.

In the following, let $k := |V_2|^2$. In order to have a subsumption relationship between P_k and F, both must "live" in the same TBox. For this, we simply take the union $\mathcal{T}_3$ of $\mathcal{T}_2^{(k)}$ and $\mathcal{T}_2$. Note that we may assume without loss of generality that the only defined concepts that $\mathcal{T}_2^{(k)}$ and $\mathcal{T}_2$ have in common are the ones from $\mathcal{T}_1$. In fact, none of the new defined concepts in $\mathcal{T}_2^{(k)}$ (i.e., the elements of $\mathcal{P}_k$) lies on a cycle, and thus we can rename them without changing the meaning of these concepts. (Note that the characterization of subsumption given in Theorem 19 implies that only for defined concepts occurring on cycles their actual names are relevant.) Thus, $\mathcal{T}_3$ is a conservative extension of both $\mathcal{T}_2^{(k)}$ and $\mathcal{T}_2$.

Lemma 30 $(A, B)^0 \sqsubseteq_{\mathcal{T}_3} F$

Proof. We must show that there is an $(F, (A, B)^0)$-synchronized simulation relation $Y \colon \mathcal{G}_{\mathcal{T}_3} \mathrel{\vec{\sim}} \mathcal{G}_{\mathcal{T}_3}$ such that $(F, (A, B)^0) \in Y$.

Again, Y is based on the "product" of Y_1 and Y_2:

$$
\begin{aligned}
Y := \{(u, p) \mid {} & (u, v_1) \in Y_1 \text{ and } (u, v_2) \in Y_2 \\
& \text{where } (v_1, v_2) \text{ is the node of } p \in \mathcal{P}_k \ \} \quad \cup \\
& \{(u, v) \mid v \in N_{def} \text{ and } (u, v) \in Y_1\}.
\end{aligned}
$$

By definition of Y, $(F, A) \in Y_1$ and $(F, B) \in Y_2$ imply $(F, (A, B)^0) \in Y$. In order to show that Y is $(F, (A, B)^0)$-synchronized, we must define an appropriate selection function S. Thus, consider the following partial $(F, (A, B)^0)$-simulation chain:

$$\begin{array}{ccccccccccc} F & = & F_0 & \stackrel{r_1}{\to} & F_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & F_{n-1} & \stackrel{r_n}{\to} & F_n \\ & & Y\!\downarrow & & Y\!\downarrow & & & & Y\!\downarrow & & \\ (A,B)^0 & = & w_0 & \stackrel{r_1}{\to} & w_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & w_{n-1} & & \end{array}$$

Since $\mathcal{T}_3$ is a conservative extension of $\mathcal{T}_2^{(k)}$, the nodes w_i are all nodes of $\mathcal{G}_2^{(k)}$, i.e., elements of $\mathcal{P}_k$ or of N_{def}.

First, assume that $w_{n-1} \in N_{def}$. But then $(F_{n-1}, w_{n-1}) \in Y_1$ by the definition of Y, and the selection function S_1 yields a node $w_n \in V_3$ such that $(w_{n-1}, r_n, w_n) \in E_3$ and $(F_n, w_n) \in Y_1$. Since $w_{n-1} \in N_{def}$ and $\mathcal{T}_3$ is a conservative extension of $\mathcal{T}_1$, $(w_{n-1}, r_n, w_n) \in E_3$ implies $w_n \in N_{def}$. Consequently, $(F_n, w_n) \in Y_1$ also yields $(F_n, w_n) \in Y$. Thus, the selection function S simply chooses w_n.

Now, assume that w_{n-1} belongs to $\mathcal{P}_k$ (and thus also the other nodes w_i). Consequently, the above partial $(F, (A, B)^0)$-simulation chain is of the following form:

$$\begin{array}{ccccccccccc} F & = & F_0 & \stackrel{r_1}{\to} & F_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & F_{n-1} & \stackrel{r_n}{\to} & F_n \\ & & Y\!\downarrow & & Y\!\downarrow & & & & Y\!\downarrow & & \\ (A,B)^0 & = & p_0 & \stackrel{r_1}{\to} & p_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & p_{n-1} & & \end{array}$$

for elements $p_1, \ldots, p_{n-1}$ of $\mathcal{P}_k$. Assume that (u_i, v_i) is the node of p_i ($i = 0, \ldots, n-1$).[1] By the definitions of $\mathcal{P}_k$, Y and E_2, this implies

- $n - 1 \leq k$,
- $u_i \neq v_i$ for $i = 0, \ldots, n-1$,
- $(F_i, u_i) \in Y_1$ and $(F_i, v_i) \in Y_2$ for $i = 0, \ldots, n-1$, and
- $(u_{i-1}, r_i, u_i) \in E_1 \subseteq E_3$ and $(v_{i-1}, r_i, v_i) \in E_1 \subseteq E_3$ for $i = 1, \ldots, n-1$.

This yields the following partial simulation chains:

$$(*) \quad \begin{array}{ccccccccccc} F & = & F_0 & \stackrel{r_1}{\to} & F_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & F_{n-1} & \stackrel{r_n}{\to} & F_n \\ & & Y_1\!\downarrow & & Y_1\!\downarrow & & & & Y_1\!\downarrow & & \\ A & = & u_0 & \stackrel{r_1}{\to} & u_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & u_{n-1} & & \end{array}$$

$$(**) \quad \begin{array}{ccccccccccc} F & = & F_0 & \stackrel{r_1}{\to} & F_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & F_{n-1} & \stackrel{r_n}{\to} & F_n \\ & & Y_2\!\downarrow & & Y_2\!\downarrow & & & & Y_2\!\downarrow & & \\ B & = & v_0 & \stackrel{r_1}{\to} & v_1 & \stackrel{r_2}{\to} & \cdots & \stackrel{r_{n-1}}{\to} & v_{n-1} & & \end{array}$$

[1] The level of p_i is obviously i.

The selection functions S_1, S_2 thus yield nodes u_n, v_n such that

- $(u_{n-1}, r_n, u_n) \in E_1 \subseteq E_3$ and $(F_n, u_n) \in Y_1$;
- $(v_{n-1}, r_n, v_n) \in E_1 \subseteq E_3$ and $(F_n, v_n) \in Y_2$.

Case 1: $u_n = v_n$.
In this case, $(p_{n-1}, r_n, u_n) \in E_2^{(k)} \subseteq E_3$, and $(F_n, u_n) \in Y$. Thus, the selection function can choose u_n.

Case 2: $u_n \neq v_n$.
We will show that this implies $n \leq k$. Consequently, $p_n := (u_n, v_n)^n \in \mathcal{P}_k$, and thus $(p_{n-1}, r_n, p_n) \in E_{\mathcal{P}_k} \subseteq E_2^{(k)} \subseteq E_3$ and $(F_n, p_n) \in Y$. Hence, the selection function can choose p_n.

Assume to the contrary that $n > k$. Consider the partial simulation chains $(*)$ and $(**)$ from above. Since $k = |V_2|^2$ and $n-1 \geq k$, there exist indices $m_1, m_2 \leq n-1$ such that $F_{m_1} = u_{m_1}$ and $F_{m_2} = v_{m_2}$ (by Lemma 27). However, since the selection functions S_1, S_2 were assumed to be nice, we have $F_m = u_m$ for all $m \geq m_1$ and $F_{m'} = v_{m'}$ for all $m' \geq m_2$. Consequently, $u_{n-1} = F_{n-1} = v_{n-1}$, which contradicts our assumption that (u_{n-1}, v_{n-1}) is the node of the element p_{n-1} of $\mathcal{P}_k$.

Why does S satisfy the synchronization property? Since the directed acyclic graph $(\mathcal{P}_k, E_{\mathcal{P}_k})$ only contains paths of length $\leq k$, any infinite $(F, (A, B)^0)$-simulation chain can only have finitely many elements of $\mathcal{P}_k$ in the lower component. After that, the lower component only contains elements from N_{def}. Restricted to these nodes in the second component, Y coincides with Y_1. Since Y_1 satisfies the synchronization property, this implies that Y satisfies this property as well. □

In the following, we assume without loss of generality that the TBoxes $\mathcal{T}_2^{(k)}$ $(k \geq 0)$ are renamed such that they share only the defined concepts of $\mathcal{T}_1$. For example, in addition to the upper index describing the level of a node in $\mathcal{P}_k$ we could add a lower index k. Thus, $(u, v)_k^n$ denotes a node on level n in $\mathcal{P}_k$. For $k \geq 0$, we denote $(A, B)_k^0$ by P_k. Using this notation, we can reformulate what we have shown until now as follows: every P_k is a common subsumer of A, B, and if F is a common subsumer of A, B then there is a k such that F subsumes P_k.

As a consequence of this lemma we can show that an lcs of A, B must be equivalent to one of the P_k.

Theorem 31 *Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox and A, B defined concepts in $\mathcal{T}_1$. Then A, B in $\mathcal{T}_1$ have an lcs iff there is a $k \geq 0$ such that P_k in $\mathcal{T}_2^{(k)}$ is the lcs of A, B in $\mathcal{T}_1$.*

Proof. The direction from right to left is trivial. Thus, assume that $\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$ and that P in $\mathcal{T}_2$ is the lcs of A, B.

We define $k := n^2$ where n is the number of defined concepts in $\mathcal{T}_2$. Let $\mathcal{T}_3$ be the union of $\mathcal{T}_2$ and $\mathcal{T}_2^{(k)}$, where we assume without loss of generality that the only defined concepts shared by $\mathcal{T}_2$ and $\mathcal{T}_2^{(k)}$ are the ones in $\mathcal{T}_1$. Then Lemma 30 shows that $P_k \sqsubseteq_{\mathcal{T}_3} P$.

Since P_k is a common subsumer of A, B, the fact that P is the least common subsumer of A, B implies that subsumption in the other direction holds as well: $P \sqsubseteq_{\mathcal{T}_3} P_k$. Thus, P and P_k are equivalent, and this implies that P_k is also an lcs of A, B. □

The concepts P_k form a decreasing chain w.r.t. subsumption.

Lemma 32 *Let $\mathcal{T} := \mathcal{T}_2^{(k)} \cup \mathcal{T}_2^{(k+1)}$. Then $P_{k+1} \sqsubseteq_{\mathcal{T}} P_k$.*

Proof. First note that $\mathcal{T}$ is a conservative extension of both $\mathcal{T}_2^{(k)}$ and $\mathcal{T}_2^{(k+1)}$.

The simulation relation Z with $(P_k, P_{k+1}) \in Z$ is defined as follows:

$$Z := \{((u,v)_k^n, (u,v)_{k+1}^n) \mid (u,v)_k^n \in \mathcal{P}_k\} \cup \{(u,u) \mid u \in N_{def}\}.$$

It is easy to see that Z is indeed a synchronized simulation relation. □

The concepts A, B have an lcs iff this decreasing chain becomes stable.

Corollary 33 *P_k is the lcs of A, B iff it is equivalent to P_{k+i} for all $i \geq 1$.*

Proof. In this proof we do not explicitly name the TBoxes w.r.t. whom the subsumption relationships hold. Basically, these TBoxes are all conservative extensions of $\mathcal{T}_1$ obtained as union with some of the TBoxes $\mathcal{T}_2^{(\ell)}$. Since these TBoxes share only the defined concepts in $\mathcal{T}_1$ and the names of their newly defined concepts are irrelevant for subsumption, it is always possible to choose the right extension.

($\Rightarrow$) Lemma 32 implies that P_k subsumes P_{k+i}. Since P_{k+i} is a common subsumer of A, B, the fact that P_k is the lcs of A, B implies that P_{k+i} also subsumes P_k.

($\Leftarrow$) We know that P_k is a common subsumer of A, B. It remains to be shown that it is the *least* common subsumer. Thus, assume that F is a common subsumer of A, B. We must show that F subsumes P_k.

By Lemma 30 there is an ℓ such that F subsumes P_ℓ. If $\ell \leq k$, then Lemma 32 implies that P_ℓ subsumes P_k, and thus F subsumes P_k. If $\ell > k$, then our assumption (right-hand side of the corollary) yields that P_k and P_ℓ are equivalent, and this again implies that F subsumes P_k. □

Example 34 Let us reconsider the TBox $\mathcal{T}_1$ defined in Theorem 24. In this case, the TBoxes $\mathcal{T}_2^{(k)}$ are basically of the form[2]

$$\mathcal{T}_1 \cup \{P_k \equiv \exists r.(A,B)_k^1,\ (A,B)_k^1 \equiv \exists r.(A,B)_k^2,\ \ldots,\ (A,B)_k^{k-1} \equiv \exists r.(A,B)_k^k\},$$

and it is easy to see that there always is a strict subsumption relationship between P_k and P_{k+1} (since P_{k+1} requires an r-chain of length $k+1$ whereas P_k only requires one of length k).

The following is an example where the lcs exists.

Example 35 Let us consider the following TBox

$$\mathcal{T}_1 := \{A \equiv \exists r.A \sqcap \exists r.C,\ B \equiv \exists r.B \sqcap \exists r.C,\ C \equiv \exists r.C\}.$$

In this case, $k = 0$ does the job, and thus the lcs of A, B is P_0:

$$\mathcal{T}_2^{(0)} := \mathcal{T}_1 \cup \{P_0 \equiv \exists r.C\}.$$

In fact, it is easy to see that the path $P_0 \xrightarrow{r} C \xrightarrow{r} C \xrightarrow{r} \cdots$ can simulate any path starting with some P_ℓ for $\ell \geq 1$. Since the infinite paths starting with P_ℓ must eventually also lead to C (after at most ℓ steps), this really yields a synchronized simulation relation.

The next example is very similar to the previous one. However, in this case the lcs does not exist.

Example 36 Let us consider the following TBox

$$\mathcal{T}_1 := \{A \equiv \exists r.A \sqcap \exists r.C \sqcap \exists r.D,\ B \equiv \exists r.B \sqcap \exists r.C \sqcap \exists r.D,\ C \equiv \exists r.C,\ D \equiv \exists r.D\}.$$

In this case, there always is a strict subsumption relationship between P_k and P_{k+1} for the following reason. Consider the path

$$P_{k+1} \xrightarrow{r} (A,B)_{k+1}^1 \xrightarrow{r} \cdots \xrightarrow{r} (A,B)_{k+1}^{k+1}$$

issuing from P_{k+1}. If this path is simulated by a path

$$P_k \xrightarrow{r} u_1 \xrightarrow{r} \cdots \xrightarrow{r} u_{k+1}$$

issuing from P_k, then either $u_{k+1} = C$ or $u_{k+1} = D$. Assume without loss of generality that $u_{k+1} = C$. Then we cannot get synchronization when simulating the path

$$P_{k+1} \xrightarrow{r} (A,B)_{k+1}^1 \xrightarrow{r} \cdots \xrightarrow{r} (A,B)_{k+1}^{k+1} \xrightarrow{r} D.$$

[2] We have restricted the attention to elements of $\mathcal{P}_k$ that are reachable from P_k.

One might think that the decreasing chain of concepts P_k ($k \geq 0$) becomes stable as soon as P_k is equivalent to P_{k+1}. Our final example shows that this is not the case. It demonstrates that $P_k \equiv P_{k+1}$ need not imply $P_{k+1} \equiv P_{k+2}$.

Example 37 Let us consider the following TBox

$$\begin{aligned}\mathcal{T}_1 := \{ & A_1 \equiv \exists r_1.A_2, A_2 \equiv \exists r_2.A_1 \sqcap \exists r_2.C, \\ & B_1 \equiv \exists r_1.B_2, B_2 \equiv \exists r_2.B_1 \sqcap \exists r_2.C, \\ & C \equiv \exists r_2.C\}.\end{aligned}$$

First, we claim that P_1 is subsumed by P_2, and thus P_1 and P_2 are equivalent. In fact, the only critical simulation chains are those of length 2 where the nodes of the upper component all belong to $\mathcal{P}_2$:

$$(1)\quad \begin{array}{ccccccc} P_2 & = & (A_1, B_1)_2^0 & \stackrel{r_1}{\to} & (A_2, B_2)_2^1 & \stackrel{r_2}{\to} & (A_1, B_1)_2^2 \\ & & \downarrow & & \downarrow & & \downarrow \\ P_1 & = & (A_1, B_1)_1^0 & \stackrel{r_1}{\to} & (A_2, B_2)_1^1 & \stackrel{r_2}{\to} & C \end{array}$$

$$(2)\quad \begin{array}{ccccccc} P_2 & = & (A_1, B_1)_2^0 & \stackrel{r_1}{\to} & (A_2, B_2)_2^1 & \stackrel{r_2}{\to} & (A_1, C)_2^2 \\ & & \downarrow & & \downarrow & & \downarrow \\ P_1 & = & (A_1, B_1)_1^0 & \stackrel{r_1}{\to} & (A_2, B_2)_1^1 & \stackrel{r_2}{\to} & C \end{array}$$

$$(3)\quad \begin{array}{ccccccc} P_2 & = & (A_1, B_1)_2^0 & \stackrel{r_1}{\to} & (A_2, B_2)_2^1 & \stackrel{r_2}{\to} & (C, B_1)_2^2 \\ & & \downarrow & & \downarrow & & \downarrow \\ P_1 & = & (A_1, B_1)_1^0 & \stackrel{r_1}{\to} & (A_2, B_2)_1^1 & \stackrel{r_2}{\to} & C \end{array}$$

In all three cases, the upper node does not have any successor node in $\mathcal{G}_{\mathcal{T}_2^{(2)}}$, and thus these chains are unproblematic.
In contrast, P_2 is *not* subsumed by P_3. In fact, consider the following situation:

$$\begin{array}{ccccccccc} P_3 & = & (A_1, B_1)_3^0 & \stackrel{r_1}{\to} & (A_2, B_2)_3^1 & \stackrel{r_2}{\to} & (A_1, B_1)_3^2 & \stackrel{r_1}{\to} & (A_2, B_2)_3^3 \\ & & \downarrow & & \downarrow & & & & \\ P_2 & = & (A_1, B_1)_2^0 & \stackrel{r_1}{\to} & (A_2, B_2)_2^1 & & & & \end{array}$$

We can simulate $(A_1, B_1)_3^2$ only by one of the following nodes: $(A_1, B_1)_2^2$, $(A_1, C)_2^2$, $(C, B_1)_2^2$, or C. However, none of these nodes has an r_1-successor in $\mathcal{G}_{\mathcal{T}_2^{(2)}}$.

4.2.3 A sufficient condition for the existence of the lcs

If we want to use the results from the previous subsection to compute the lcs, we must be able to decide whether there is an index k such that P_k is the lcs of A, B, and if yes we must also be able to compute such a k. Though we strongly conjecture that this is possible, we have not yet found such a procedure. For this reason, we must restrict ourself to give a *sufficient condition* for the lcs of two concepts defined in an $\mathcal{EL}$-TBox to exist.

As before, let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox, let $\mathcal{G}_{\mathcal{T}_1} = (N_{def}, E_{\mathcal{T}_1}, L_{\mathcal{T}_1})$ be the corresponding description graph, and let A, B be defined concepts in $\mathcal{T}_1$ (i.e., elements of N_{def}). We consider the product $\mathcal{G} := \mathcal{G}_{\mathcal{T}_1} \times \mathcal{G}_{\mathcal{T}_1}$ of $\mathcal{G}_{\mathcal{T}_1}$ with itself. Let $\mathcal{G} = (V, E, L)$.

Definition 38 We say that (A, B) is synchronized in $\mathcal{T}_1$ iff, for every infinite path

$$(A, B) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots$$

in $\mathcal{G}$, there exists an index $i \geq 0$, such that $u_i = v_i$.

For example, in the TBox $\mathcal{T}_1$ introduced in Theorem 24, (A, B) is *not* synchronized. The same is true for the TBox defined in Example 35. As another example, consider the TBox

$$\mathcal{T}_1' := \{A' \equiv \exists r_1.A' \sqcap \exists r.C, \quad B' \equiv \exists r_2.B' \sqcap \exists r.C, \quad C \equiv \exists r.C\}.$$

In this TBox, (A', B') is synchronized.

Lemma 39 *Assume that (A, B) is synchronized in $\mathcal{T}_1$, and let $k := |N_{def}|^2$. Then, for every path*

$$(A, B) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots \xrightarrow{r_k} (u_k, v_k)$$

in $\mathcal{G}$ of length k, there exists an index $i, 0 \leq i \leq k$ such that $u_i = v_i$.

Proof. Assume to the contrary that there is a path

$$(A, B) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots \xrightarrow{r_k} (u_k, v_k)$$

such that $u_i \neq v_i$ for all $i, 0 \leq i \leq k$. Since $(u_i, v_i) \in N_{def} \times N_{def}$ for $i = 0, \ldots, k$ and $k = |N_{def} \times N_{def}|$, there exist indices $0 \leq i < j \leq k$ such that $(u_i, v_i) = (u_j, v_j)$.

But then we can construct an infinite path

$$(A, B) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} \cdots (u_i, v_i) \xrightarrow{r_{i+1}} \cdots \xrightarrow{r_j} (u_j, v_j) = (u_i, v_i) \xrightarrow{r_{i+1}} \cdots$$

such that the first component in the tuples is always different from the second component. This contradicts our assumption that (A, B) is synchronized in $\mathcal{G}$. $\square$

As an easy consequence of this lemma we obtain that $k = |N_{def}|^2$ is such that P_k is the lcs of A, B.

Lemma 40 *Assume that (A, B) is synchronized in $\mathcal{T}_1$, and let $k := |N_{def}|^2$. Then P_k in $\mathcal{T}_2^{(k)}$ is the lcs of A, B in $\mathcal{T}_1$.*

Proof. By the previous lemma, every path in $\mathcal{G}$ starting with (A, B) and of length at least k contains some node of the form (u, u). Thus, if we consider a path starting with P_ℓ for some $\ell \geq k$, then we know that only an initial segment of length $\leq k$ can belong to $\mathcal{P}_\ell$. Basically, this initial segment also belongs to $\mathcal{P}_k$ (modulo the lower index). This observation can be used to show that P_k is equivalent to P_ℓ for all $\ell \geq k$, and thus P_k is the lcs of A, B. $\square$

As an immediate consequence of Lemma 40 we obtain that the lcs of A, B in $\mathcal{T}_1$ always exists, provided that (A, B) is synchronized in $\mathcal{T}_1$. Our construction of the TBox $\mathcal{T}_2^{(k)}$ is obviously polynomial in k and the size of $\mathcal{T}_1$. Since k is also polynomial in the size of $\mathcal{T}_1$, the size of $\mathcal{T}_2$ is polynomial in the size of $\mathcal{T}_1$.

Theorem 41 *Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox, and let A, B be defined concepts in $\mathcal{T}_1$ such that (A, B) is synchronized in $\mathcal{T}_1$. Then the lcs of A, B in $\mathcal{T}_1$ always exists, and it can be computed in polynomial time.*

Example 24 shows that the lcs may exist even if (A, B) is not synchronized in $\mathcal{T}_1$. Thus, this is a sufficient, but not necessary condition for the existence of the lcs.

We close this section by showing that this sufficient condition can be decided in polynomial time.

Proposition 42 *Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox, and let A, B be defined concepts in $\mathcal{T}_1$. Then it can be decided in polynomial time whether (A, B) is synchronized in $\mathcal{T}_1$ or not.*

Proof. As before, consider the product $\mathcal{G} := \mathcal{G}_{\mathcal{T}_1} \times \mathcal{G}_{\mathcal{T}_1}$ of $\mathcal{G}_{\mathcal{T}_1}$ with itself. Let $\mathcal{G} = (V, E, L)$.

We define

$$
\begin{aligned}
W_0 &:= \{(u,u) \mid (u,u) \in V\}, \\
W_{i+1} &:= W_i \cup \{(u,v) \mid (u,v) \in V \text{ and all edges with source } (u,v) \text{ in } \mathcal{G} \\
&\qquad \text{lead to elements of } W_i\}, \quad \text{and} \\
W_\infty &:= \bigcup_{i \geq 0} W_i.
\end{aligned}
$$

Obviously, W_∞ can be computed in time polynomial in the size of $\mathcal{G}$.

Claim 1: (A, B) *is synchronized in* $\mathcal{T}_1$ *iff* $(A, B) \in W_\infty$.

From this, the proposition immediately follows. To prove this claim, we show the following:

Claim 2: $(u, v) \in W_n$ *iff for every path*

$$(u, v) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots \xrightarrow{r_n} (u_n, v_n)$$

in $\mathcal{G}$ of length n, there exists an index $i, 0 \leq i \leq n$ such that $u_i = v_i$.

We prove Claim 2 by induction on n. If $n = 0$, then $(u, v) \in W_n$ iff $u = v$. In addition, the above path has length 0, i.e., consists of $(u, v) = (u_0, v_0)$ only. Thus, the existence of an $i, 0 \leq i \leq 0$ such that $u_i = v_i$ is equivalent to $u = v$.

$(n \to n + 1)$ First, assume that $(u, v) \in W_{n+1}$. If $(u, v) \in W_n$, then the induction hypothesis can be applied. Thus, assume that $(u, v) \in W_{n+1} \setminus W_n$, i.e., all edges with source (u, v) in $\mathcal{G}$ lead to elements of W_n. Now, consider a path

$$(u, v) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots \xrightarrow{r_n} (u_n, v_n) \xrightarrow{r_{n+1}} (u_{n+1}, v_{n+1})$$

in $\mathcal{G}$ of length $n + 1$. Since $(u_1, v_1) \in W_n$, there exists an index $i, 1 \leq i \leq n + 1$ such that $u_i = v_i$, and we are done.

Second assume that, for every path

$$(*) \quad (u, v) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots \xrightarrow{r_n} (u_n, v_n) \xrightarrow{r_{n+1}} (u_{n+1}, v_{n+1})$$

in $\mathcal{G}$ of length $n + 1$, there exists an index $i, 0 \leq i \leq n + 1$ such that $u_i = v_i$. If $u = v$, then $(u, v) \in W_0 \subseteq W_{n+1}$. Thus, assume that $u \neq v$. To show that $(u, v) \in W_{n+1}$, we consider an arbitrary edge $((u, v), r_1, (u_1, v_1))$ in $\mathcal{G}$ and show that $(u_1, v_1) \in W_n$. Thus, consider a path in $\mathcal{G}$ of length n starting with (u_1, v_1):

$$(u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots \xrightarrow{r_n} (u_n, v_n) \xrightarrow{r_{n+1}} (u_{n+1}, v_{n+1}).$$

Together with the edge $((u, v), r_1, (u_1, v_1))$ this yields a path of length $n + 1$ of the form $(*)$ above. Thus, there exists an index $i, 0 \leq i \leq n+1$ such that $u_i = v_i$. Since we have assumed that $u \neq v$, we actually have $1 \leq i \leq n + 1$, which shows that $(u_1, v_1) \in W_n$. This completes the proof of Claim 2.

It remains to be shown that Claim 2 implies Claim 1. First, assume that (A, B) is synchronized in $\mathcal{T}_1$. By Lemma 39, there is a k such that, for every path

$$(u, v) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots \xrightarrow{r_k} (u_k, v_k)$$

in $\mathcal{G}$ of length k, there exists an index $i, 0 \leq i \leq k$ such that $u_i = v_i$. By Claim 2, this implies $(A, B) \in W_k \subseteq W_\infty$.

Conversely, assume that $(A, B) \in W_\infty$. Thus, there is a k such that $(A, B) \in W_k$. By Claim 2, this implies that, for every path

$$(u, v) = (u_0, v_0) \xrightarrow{r_1} (u_1, v_1) \xrightarrow{r_2} (u_2, v_2) \xrightarrow{r_3} \cdots \xrightarrow{r_k} (u_k, v_k)$$

in $\mathcal{G}$ of length k, there exists an index $i, 0 \leq i \leq k$ such that $u_i = v_i$. In particular, this shows that (A, B) is synchronized in $\mathcal{T}_1$. $\square$

5 The instance problem and most specific concepts

One motivation for considering cyclic terminologies in $\mathcal{EL}$ is the fact that the most specific concept of an ABox individual need not exist in $\mathcal{EL}$ (without cyclic terminologies). An example is the simple cyclic ABox $\mathcal{A} := \{r(b,b)\}$, where b has no most specific concept, i.e., there is no least $\mathcal{EL}$-concept description D such that b is an instance of D w.r.t. $\mathcal{A}$ [9]. However, if one allows for cyclic TBoxes with gfp-semantics, then the defined concept B with $B \equiv \exists r.B$ is such a most specific concept.

In the following, we restrict the attention to gfp-semantics. First, we show how the characterization of subsumption (Theorem 13) can be extended to the instance problem w.r.t. gfp-semantics. Then, we will use this characterization to characterize the most specific concept w.r.t. gfp-semantics (gfp-msc).

5.1 The instance problem w.r.t. gfp-semantics

Assume that $\mathcal{T}$ is an $\mathcal{EL}$-TBox and $\mathcal{A}$ an $\mathcal{EL}$-ABox. In the following, we assume that $\mathcal{T}$ is fixed and that all instance problems for $\mathcal{A}$ are considered w.r.t. this TBox.

In this setting, $\mathcal{A}$ can be translated into an $\mathcal{EL}$-description graph $\mathcal{G}_\mathcal{A}$ by viewing $\mathcal{A}$ as a graph and extending it appropriately by the graph $\mathcal{G}_\mathcal{T}$ associated with $\mathcal{T}$. The idea is then that the characterization of the instance problem should be similar to the statement of Proposition 15: the individual a is an instance of A in $\mathcal{A}$ iff there is a simulation $Z\colon \mathcal{G}_\mathcal{T} \overset{\rightarrow}{\sim} \mathcal{G}_\mathcal{A}$ such that $(A,a) \in Z$.

Before giving an exact definition of $\mathcal{G}_\mathcal{A}$, we consider an example that demonstrates that a too simple-minded realization of this idea does not work. Let

$$\begin{array}{rcl} \mathcal{A} & := & \{A(a), P(a)\} \text{ and} \\ \mathcal{T} & := & \{A \equiv \exists r.A,\ B \equiv P \sqcap \exists r.B\}. \end{array}$$

The ABox $\mathcal{A}$ itself can be viewed as an $\mathcal{EL}$-description graph consisting of a single node a with label $\{P\}$. Since $A \equiv \exists r.A$ is in $\mathcal{T}$ and since $A(a)$ is in $\mathcal{A}$, we extend this graph by an r-loop from a to a. Figure 3 shows the graph $\mathcal{G}$ obtained this way as well as the $\mathcal{EL}$-description graph $\mathcal{G}_\mathcal{T}$ corresponding to $\mathcal{T}$.

Obviously, there is a simulation $Z\colon \mathcal{G}_\mathcal{T} \overset{\rightarrow}{\sim} \mathcal{G}$ such that $(B,a) \in Z$. However, a is not an instance of B. In fact, of a we only know that it belongs to P and that it is the starting point of an infinite r-chain. The instances of B must belong to P and they must be the starting point of an infinite r-chain such that every element on this chain also belongs to P.

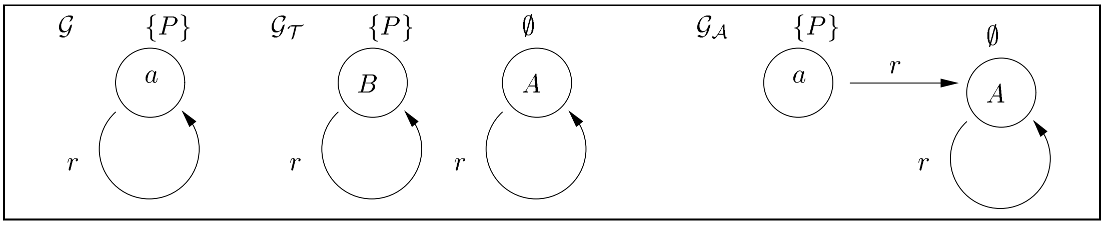

Figure 3: The $\mathcal{EL}$-description graphs $\mathcal{G}$, $\mathcal{G}_\mathcal{T}$, and $\mathcal{G}_\mathcal{A}$ of our example.

The reason for this problem is that our construction of $\mathcal{G}$ was too simple minded. In fact, node labels and edges in $\mathcal{G}_\mathcal{T}$ state facts that must hold for all individuals that are instances of the defined concept labeling a given node. Assertions of the ABox make statements about properties of particular named individuals. The construction of $\mathcal{G}$ in the above example mixes these different things, and thus leads to unfounded conclusions.

In order to separate edges and labels coming from ABox assertions from the ones coming from TBox definitions, we do not "identify" the node a with the node A if $A(a)$ belongs to $\mathcal{A}$ (as done in the construction of $\mathcal{G}$ above). Instead, we do a "one-step expansion" of the definition of A. The right-most graph in Figure 3 shows the graph $\mathcal{G}_\mathcal{A}$ obtained this way in our example. Obviously, there is no simulation Z: $\mathcal{G}_\mathcal{T} \overset{\rightarrow}{\sim} \mathcal{G}_\mathcal{A}$ such that $(B, a) \in Z$.

Below, we give a formal definition of the $\mathcal{EL}$-description graph $\mathcal{G}_\mathcal{A}$ associated with the ABox $\mathcal{A}$ and the TBox $\mathcal{T}$ in the general case.

Definition 43 Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{A}$ an $\mathcal{EL}$-ABox, and $\mathcal{G}_\mathcal{T} = (V, E, L)$ be the $\mathcal{EL}$-description graph associated with $\mathcal{T}$. The $\mathcal{EL}$-description graph $\mathcal{G}_\mathcal{A} = (V_\mathcal{A}, E_\mathcal{A}, L_\mathcal{A})$ associated with $\mathcal{A}$ and $\mathcal{T}$ is defined as follows:

- the nodes of $\mathcal{G}_\mathcal{A}$ are the individual names occurring in $\mathcal{A}$ together with the defined concepts of $\mathcal{T}$, i.e.,

$$V_\mathcal{A} := V \cup \{a \mid a \text{ is an individual name occurring in } \mathcal{A}\};$$

- the edges of $\mathcal{G}_\mathcal{A}$ are the edges of $\mathcal{G}$, the role assertions of $\mathcal{A}$, and additional edges linking the ABox individuals with defined concepts:

$$\begin{array}{rcl} E_\mathcal{A} & := & E \cup \{(a, r, b) \mid r(a, b) \in \mathcal{A}\} \ \cup \\ & & \{(a, r, B) \mid A(a) \in \mathcal{A} \text{ and } (A, r, B) \in E\}; \end{array}$$

- if $u \in V_\mathcal{A}$ is a defined concept, then it inherits its label from $\mathcal{G}_\mathcal{T}$, i.e.,

$$L_\mathcal{A}(u) := L(u) \quad \text{if } u \in V;$$

otherwise, u is an ABox individual, and then its label is derived from the concept assertions for u in $\mathcal{A}$. In the following, let P denote primitive and A denote defined concepts.

$$L_{\mathcal{A}}(u) := \{P \mid P(u) \in \mathcal{A}\} \cup \bigcup_{A(u)\in\mathcal{A}} L(A) \quad \text{if } u \in V_{\mathcal{A}} \setminus V.$$

Before we can characterize the instance problem via the existence of certain simulation relations from $\mathcal{G}_{\mathcal{T}}$ to $\mathcal{G}_{\mathcal{A}}$, we must characterize under what conditions a gfp-model of $\mathcal{T}$ is a model of $\mathcal{A}$. In the following we assume that primitive interpretations also interpret ABox individuals.

Definition 44 Let $\mathcal{J}$ be a primitive interpretation and $\mathcal{G}_{\mathcal{J}}$ the $\mathcal{EL}$-description graph associated with $\mathcal{J}$. We say that the simulation $Z\colon \mathcal{G}_{\mathcal{A}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{J}}$ *respects ABox individuals* iff

$$\{x \mid (a, x) \in Z\} = \{a^{\mathcal{J}}\}$$

for all individual names a occurring in $\mathcal{A}$.

Proposition 45 *Let $\mathcal{J}$ be a primitive interpretation and $\mathcal{I}$ the gfp-model of $\mathcal{T}$ based on $\mathcal{J}$. Then the following are equivalent:*

1. *$\mathcal{I}$ is a model of $\mathcal{A}$.*
2. *There is a simulation $Z\colon \mathcal{G}_{\mathcal{A}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{J}}$ that respects ABox individuals.*

Proof. $(2 \Rightarrow 1)$ Assume that a simulation $Z\colon \mathcal{G}_{\mathcal{A}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{J}}$ respecting ABox individuals is given. We must show that $\mathcal{I}$ satisfies all the assertions in $\mathcal{A}$.

First, assume that $r(a, b)$ is a role assertion in $\mathcal{A}$. Since $\mathcal{I}$ coincides with $\mathcal{J}$ on role and individual names, we must show that $(a^{\mathcal{J}}, b^{\mathcal{J}}) \in r^{\mathcal{J}}$. Because Z respects ABox individuals, we know that $(a, a^{\mathcal{J}}) \in Z$, and thus $(a, r, b) \in E_{\mathcal{A}}$ implies that there exists a $y \in \Delta^{\mathcal{J}}$ such that $(a^{\mathcal{J}}, y) \in r^{\mathcal{J}}$ and $(b, y) \in Z$. Since Z respects ABox individuals, $(b, y) \in Z$ implies that $y = b^{\mathcal{J}}$, which yields $(a^{\mathcal{J}}, b^{\mathcal{J}}) \in r^{\mathcal{J}}$.

Second, assume that $P(a)$ is a concept assertion in $\mathcal{A}$ where P is a primitive concept. By definition of $L_{\mathcal{A}}$, we have $P \in L_{\mathcal{A}}(a)$. In addition, since Z respects ABox individuals, we know that $(a, a^{\mathcal{J}}) \in Z$, which implies $L_{\mathcal{A}}(a) \subseteq L_{\mathcal{J}}(a^{\mathcal{J}})$. Consequently, $P \in L_{\mathcal{J}}(a^{\mathcal{J}})$, which implies $a^{\mathcal{I}} = a^{\mathcal{J}} \in P^{\mathcal{J}} = P^{\mathcal{I}}$.

Finally, assume that $A(a)$ is a concept assertion in $\mathcal{A}$ where A is a defined concept. We use Proposition 15 to show that $a^{\mathcal{J}} = a^{\mathcal{I}} \in A^{\mathcal{I}}$. Thus, we need to find a simulation $Y\colon \mathcal{G}_{\mathcal{T}} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{J}}$ such that $(A, a^{\mathcal{J}}) \in Y$. We define the relation Y as follows:

$$Y := \{(A, a^{\mathcal{J}})\} \cup \{(B, x) \mid (B, x) \in Z \text{ and } B \in V\}.$$

Thus, Y is the restriction of Z to the nodes of $\mathcal{G}_\mathcal{T}$, extended by the tuple $(A, a^\mathcal{J})$. It remains to be shown that Y is a simulation relation, i.e., satisfies (S1) and (S2) of Definition 11.

(S1) Let $(B, x) \in Y$. If $(B, x) \in Z$, then $L(B) = L_\mathcal{A}(B) \subseteq L_\mathcal{J}(x)$ since Z is a simulation. Thus, consider $(A, a^\mathcal{J}) \in Y$ and let $P \in L(A)$. By definition of $\mathcal{G}_\mathcal{A}$, $A(a) \in \mathcal{A}$ and $P \in L(A)$ imply that $P \in L_\mathcal{A}(a)$. Since Z respects ABox individuals, we know that $(a, a^\mathcal{J}) \in Z$, and thus $P \in L_\mathcal{A}(a) \subseteq L_\mathcal{J}(a^\mathcal{J})$.

(S2) Let $(B, x) \in Y$ and $(B, r, B') \in E$. If $(B, x) \in Z$, then $(B, r, B') \in E \subseteq E_\mathcal{A}$ implies the existence of a y such that $(B', y) \in Z$ and $(x, r, y) \in E_\mathcal{J}$. Now, $(B, r, B') \in E$ yields $B' \in V$, and thus $(B', y) \in Z$ implies $(B', y) \in Y$.

Thus, consider $(A, a^\mathcal{J}) \in Y$ and $(A, r, B') \in E$. Since $A(a) \in \mathcal{A}$, the definition of $\mathcal{G}_\mathcal{A}$ shows that $(a, r, B') \in E_\mathcal{A}$. In addition, since Z respects ABox individuals, we know that $(a, a^\mathcal{J}) \in Z$. Consequently, there is a y such that $(a^\mathcal{J}, r, y) \in E_\mathcal{J}$ and $(B', y) \in Z$. Now, $(A, r, B') \in E$ yields $B' \in V$, and thus $(B', y) \in Y$.

This completes the proof that $Y \colon \mathcal{G}_\mathcal{T} \rightleftarrows \mathcal{G}_\mathcal{J}$ is a simulation such that $(A, a^\mathcal{J}) \in Y$. Thus, Proposition 15 implies that $a^\mathcal{I} = a^\mathcal{J} \in A^\mathcal{I}$.

$(1 \Rightarrow 2)$ Assume that $\mathcal{I}$ is a model of $\mathcal{A}$. In particular, this implies that $a^\mathcal{J} = a^\mathcal{I} \in A^\mathcal{I}$ holds for all concept assertions $A(a) \in \mathcal{A}$. Thus, Proposition 15 implies the existence of simulation relations $Z_{A(a)} \colon \mathcal{G}_\mathcal{T} \rightleftarrows \mathcal{G}_\mathcal{J}$ such that $(A, a^\mathcal{J}) \in Z_{A(a)}$. Let Y be the union of these simulations, i.e.,

$$Y := \bigcup_{A(a) \in \mathcal{A}} Z_{A(a)}.$$

Then Y is a simulation relation that satisfies $(A, a^\mathcal{J}) \in Y$ for all concept assertions $A(a) \in \mathcal{A}$. We define the relation Z as follows:

$$Z := Y \cup \{(a, a^\mathcal{J}) \mid a \text{ is an individual name occurring in } \mathcal{A}\}.$$

By definition of Z, $\{x \mid (a, x) \in Z\} = \{a^\mathcal{J}\}$, and thus it remains to be shown that Z is a simulation from $\mathcal{G}_\mathcal{A}$ to $\mathcal{G}_\mathcal{J}$.

(S1) Since Y satisfies this property, it is enough to consider the case $(a, a^\mathcal{J}) \in Z$. If $P \in L_\mathcal{A}(a)$, then $P(a) \in \mathcal{A}$ or $P \in L(A)$ and $A(a) \in \mathcal{A}$ for some defined concept A. In both cases, the fact that $\mathcal{I}$ is a model of $\mathcal{A}$ implies that $a^\mathcal{J} = a^\mathcal{I} \in P^\mathcal{I} = P^\mathcal{J}$, and thus $P \in L_\mathcal{J}(a^\mathcal{J})$.

(S2) First, consider the case where $(B, x) \in Y$ for some defined concept B and element x of $\Delta^\mathcal{J}$. Since any edge (B, r, u) in $\mathcal{G}_\mathcal{A}$ with source $B \in V$ is also an edge in $\mathcal{G}_\mathcal{T}$, the fact that Y satisfies (S2) implies the existence of a y such that $(x, r, y) \in E_\mathcal{J}$ and $(u, y) \in Y \subseteq Z$.

Second, consider $(a, a^\mathcal{J}) \in Z$ and assume that $(a, r, u) \in E_\mathcal{A}$. By the definition of $E_\mathcal{A}$, this means that one of the following two cases applies:

1. u is an ABox individual and $r(a, u) \in \mathcal{A}$.
2. u is a defined concept and $(A, r, u) \in E$ and $A(a) \in \mathcal{A}$ for some defined concept A.

In the first case, we have $(u, u^{\mathcal{J}}) \in Z$, and $(a^{\mathcal{J}}, u^{\mathcal{J}}) \in r^{\mathcal{J}}$ since $\mathcal{I}$ is a model of $\mathcal{A}$ and coincides with $\mathcal{J}$ on role and individual names. Since $(a^{\mathcal{J}}, u^{\mathcal{J}}) \in r^{\mathcal{J}}$ is equivalent to $(a^{\mathcal{J}}, r, u^{\mathcal{J}}) \in E_{\mathcal{J}}$, we have shown that (S2) holds in this case.

In the second case, $A(a) \in \mathcal{A}$ implies that $(A, a^{\mathcal{J}}) \in Y$ by the definition of Y. Since Y is a simulation, this together with $(A, r, u) \in E$ implies that there exists a y such that $(u, y) \in Y \subseteq Z$ and $(a^{\mathcal{J}}, r, y) \in E_{\mathcal{J}}$. This completes the proof of the proposition. □

The following characterization of the instance problem is an easy consequence of this proposition.

Theorem 46 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{A}$ an $\mathcal{EL}$-ABox, A a defined concept in $\mathcal{T}$ and a an individual name occurring in $\mathcal{A}$. Then the following are equivalent:*

1. *$\mathcal{A} \models_{gfp,\mathcal{T}} A(a)$.*
2. *There is a simulation Z: $\mathcal{G}_{\mathcal{T}} \rightleftarrows \mathcal{G}_{\mathcal{A}}$ such that $(A, a) \in Z$.*

Proof. ($2 \Rightarrow 1$) Assume that there is a simulation Z: $\mathcal{G}_{\mathcal{T}} \rightleftarrows \mathcal{G}_{\mathcal{A}}$ such that $(A, a) \in Z$. We must show $\mathcal{A} \models_{gfp,\mathcal{T}} A(a)$, i.e., if $\mathcal{I}$ is a gfp-model of $\mathcal{T}$ that is also a model of $\mathcal{A}$, then $a^{\mathcal{I}} \in A^{\mathcal{I}}$. Thus, let $\mathcal{J}$ be a primitive interpretation and $\mathcal{I}$ the gfp-model of $\mathcal{T}$ based on $\mathcal{J}$.

If $\mathcal{I}$ is a model of $\mathcal{A}$, then Proposition 45 yields a simulation Y: $\mathcal{G}_{\mathcal{A}} \rightleftarrows \mathcal{G}_{\mathcal{J}}$ that respects ABox individuals. The composition $X := Z \circ Y$ is a simulation from $\mathcal{G}_{\mathcal{T}}$ to $\mathcal{G}_{\mathcal{J}}$ such that $(A, a^{\mathcal{J}}) \in X$. In fact, we know that $(A, a) \in Z$ and the fact that Y respects ABox individuals implies that $(a, a^{\mathcal{J}}) \in Y$. Thus, Proposition 15 yields $a^{\mathcal{I}} = a^{\mathcal{J}} \in A^{\mathcal{I}}$.

($1 \Rightarrow 2$) Assume that $\mathcal{A} \models_{gfp,\mathcal{T}} A(a)$. The $\mathcal{EL}$-description graph $\mathcal{G}_{\mathcal{A}}$ can be viewed as the graph of a primitive interpretation. Thus, let $\mathcal{J}$ be this primitive interpretation, i.e., $\mathcal{G}_{\mathcal{A}} = \mathcal{G}_{\mathcal{J}}$, and let $\mathcal{I}$ be the gfp-model of $\mathcal{T}$ based on $\mathcal{J}$.

We claim that $\mathcal{I}$ is a model of $\mathcal{A}$. This is an immediate consequence of Proposition 45 since the identity on $\mathcal{G}_{\mathcal{A}}$ is a simulation from $\mathcal{G}_{\mathcal{A}}$ to $\mathcal{G}_{\mathcal{J}} = \mathcal{G}_{\mathcal{A}}$ that respects ABox individuals.

Consequently, the fact that $\mathcal{A} \models_{gfp,\mathcal{T}} A(a)$ implies that $a = a^{\mathcal{J}} = a^{\mathcal{I}} \in A^{\mathcal{I}}$. But then Proposition 15 yields the desired simulation Z: $\mathcal{G}_{\mathcal{T}} \rightleftarrows \mathcal{G}_{\mathcal{J}} = \mathcal{G}_{\mathcal{A}}$ such that $(A, a) \in Z$. □

The theorem together with Proposition 12 shows that the instance problem w.r.t. gfp-semantics in $\mathcal{EL}$ is tractable.

Corollary 47 *The instance problem w.r.t. gfp-semantics in $\mathcal{EL}$ can be decided in polynomial time.*

5.2 Computing the gfp-msc

Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox and $\mathcal{A}$ an $\mathcal{EL}$-ABox containing the individual name a. Let $\mathcal{G}_\mathcal{A} = (V_\mathcal{A}, E_\mathcal{A}, L_\mathcal{A})$ be the $\mathcal{EL}$-description graph corresponding to $\mathcal{A}$ and $\mathcal{T}_1$, as introduced in Definition 43. In order to obtain the gfp-msc of a, we view $\mathcal{G}_\mathcal{A}$ as the $\mathcal{EL}$-description graph of an $\mathcal{EL}$-TBox $\mathcal{T}_2$, i.e., let $\mathcal{T}_2$ be the TBox such that $\mathcal{G}_\mathcal{A} = \mathcal{G}_{\mathcal{T}_2}$. By the definition of $\mathcal{G}_\mathcal{A}$, the defined concepts of $\mathcal{T}_2$ are the defined concepts of $\mathcal{T}_1$ together with the individual names occurring in $\mathcal{A}$.

Lemma 48 *$\mathcal{T}_2$ is a conservative extension of $\mathcal{T}_1$*

Proof. This is an easy consequence of the definitions of $E_\mathcal{A}$ and $L_\mathcal{A}$. □

To avoid confusion we will denote the defined concept in $\mathcal{T}_2$ corresponding to the individual name b in $\mathcal{A}$ by C_b. Using the results of the previous subsection, we can show that C_a is the gfp-msc of a.

Proposition 49 *The defined concept C_a in $\mathcal{T}_2$ is the gfp-msc of a in $\mathcal{A}$ and $\mathcal{T}_1$.*

Proof. First, we show that a is an instance of C_a w.r.t. gfp-semantics, i.e., $\mathcal{A} \models_{gfp,\mathcal{T}_2} C_a(a)$. The identity on $\mathcal{G}_{\mathcal{T}_2}$ is a simulation from $\mathcal{G}_{\mathcal{T}_2} = \mathcal{G}_\mathcal{A}$ to $\mathcal{G}_\mathcal{A}$ that contains the tuple (C_a, a).[3] Thus, Theorem 46 yields $\mathcal{A} \models_{gfp,\mathcal{T}_2} C_a(a)$.

Second, assume that $\mathcal{T}_3$ is a conservative extension of $\mathcal{T}_2$ and that F is a defined concept in $\mathcal{T}_3$ such that $\mathcal{A} \models_{gfp,\mathcal{T}_3} F(a)$. Let $\widehat{\mathcal{G}}_\mathcal{A}$ be the $\mathcal{EL}$-description graph corresponding to $\mathcal{A}$ and $\mathcal{T}_3$, as introduced in Definition 43. By Theorem 46, $\mathcal{A} \models_{gfp,\mathcal{T}_3} F(a)$ implies that there is a simulation $Z\colon \mathcal{G}_{\mathcal{T}_3} \overset{\rightarrow}{\sim} \widehat{\mathcal{G}}_\mathcal{A}$ such that $(F, a) \in Z$. We must show that $C_a \sqsubseteq_{gfp,\mathcal{T}_3} F$. By Theorem 13, it is enough to show that there is a simulation $Y\colon \mathcal{G}_{\mathcal{T}_3} \overset{\rightarrow}{\sim} \mathcal{G}_{\mathcal{T}_3}$ such that $(F, C_a) \in Y$.

To define Y, first note that the set of nodes of $\widehat{\mathcal{G}}_\mathcal{A}$ consists of the nodes of $\mathcal{G}_{\mathcal{T}_3}$ and the individuals occurring in $\mathcal{A}$. Also note that $\mathcal{T}_3$ extends $\mathcal{T}_2$, and that $\mathcal{G}_{\mathcal{T}_2}$ in principle also contains the individuals occurring in $\mathcal{A}$. However, we assume without loss of generality that the individual names b in $\mathcal{T}_2$ have been renamed into concept names C_b. The definition of $\widehat{\mathcal{G}}_\mathcal{A}$ is illustrated by Figure 4. The

[3] Recall that C_a in $\mathcal{T}_2$ corresponds to a.

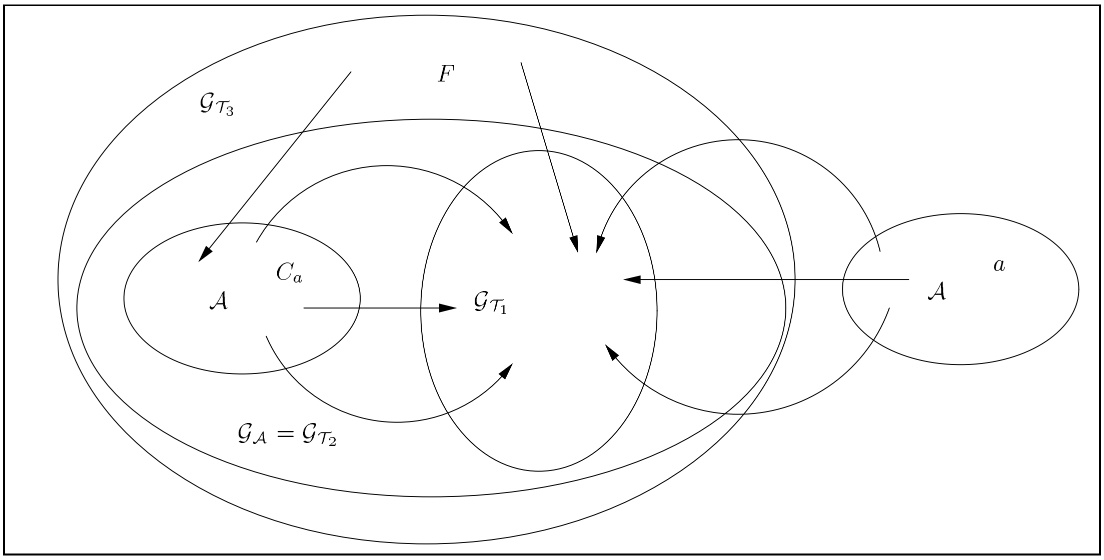

Figure 4: The $\mathcal{EL}$-description graph $\widehat{\mathcal{G}}_{\mathcal{A}}$.

arrows indicate that there may be edges from one subgraph into the other. The inner oval marked with $\mathcal{A}$ indicates the ABox $\mathcal{A}$ as used within $\mathcal{G}_{\mathcal{T}_2}$. There, the individual name a is renamed into C_a (and an analogous renaming is done for the other individual names).

The simulation Y is defined as follows:

$$\begin{aligned} Y \;\; := \;\; & \{(u,v) \in Z \mid v \text{ is a node of } \mathcal{T}_3\} \;\cup \\ & \{(u, C_b) \mid (u,b) \in Z \text{ and } b \text{ is an individual name in } \mathcal{A}\}. \end{aligned}$$

Since $(F,a) \in Z$, we have $(F, C_a) \in Y$. Thus, it remains to be shown that Y is a simulation relation.

(S1) Assume that $(u,v) \in Y$. If v is a node of $\mathcal{G}_{\mathcal{T}_3}$, then property (S1) holds since it is satisfied by Z. If $v = C_b$ for the individual name b in $\mathcal{A}$, then $(u,b) \in Z$. But then (S1) holds since the label of b in $\widehat{\mathcal{G}}_{\mathcal{A}}$ coincides with the label of C_b in $\mathcal{G}_{\mathcal{T}_2} = \mathcal{G}_{\mathcal{A}}$.

(S2) Assume that $(u,v) \in Y$ and that (u,r,u') is an edge in $\mathcal{G}_{\mathcal{T}_3}$. If v is a node of $\mathcal{G}_{\mathcal{T}_3}$, then $(u,v) \in Z$, and thus there exists a node v' such that (v,r,v') is an edge in $\widehat{\mathcal{G}}_{\mathcal{A}}$ and $(u',v') \in Z$. Since, in $\widehat{\mathcal{G}}_{\mathcal{A}}$, edges from nodes of $\mathcal{G}_{\mathcal{T}_3}$ lead to nodes of $\mathcal{G}_{\mathcal{T}_3}$, we know that v' is a node of $\mathcal{G}_{\mathcal{T}_3}$, which yields $(u',v') \in Y$.

Finally, assume that $v = C_b$ for the individual name b in $\mathcal{A}$. Then, we know that $(u,b) \in Z$, and thus there is a node v' such that (b,r,v') is an edge in $\widehat{\mathcal{G}}_{\mathcal{A}}$ and $(u',v') \in Z$. If v' is an individual name in $\mathcal{A}$, then $(u', C_{v'}) \in Y$. In addition, the existence of the edge (b,r,v') implies that there is an assertion $r(b,v') \in \mathcal{A}$. Consequently, we also have the edge $(C_b, C_{v'})$ in $\mathcal{G}_{\mathcal{T}_3}$.

It remains to consider the case where v' is a node in $\mathcal{G}_{\mathcal{T}_1}$. But then $(u', v') \in Z$ implies that $(u', v') \in Y$. In addition, the existence of the edge (b, r, v') in $\widehat{\mathcal{G}}_{\mathcal{A}}$ implies that there is the corresponding edge (C_b, v') in $\mathcal{G}_{\mathcal{A}} = \mathcal{G}_{\mathcal{T}_3}$. $\square$

Given $\mathcal{T}$ and $\mathcal{A}$, the graph $\mathcal{G}_{\mathcal{A}}$ can obviously be computed in polynomial time, and thus the gfp-msc can be computed in polynomial time.

Theorem 50 *Let $\mathcal{T}_1$ be an $\mathcal{EL}$-TBox and $\mathcal{A}$ an $\mathcal{EL}$-ABox containing the individual name a. Then the gfp-msc of a in $\mathcal{T}_1$ and $\mathcal{A}$ always exists, and it can be computed in polynomial time.*

6 Simple role-value-maps

As mentioned in the introduction, one would sometimes like to express certain relationships between roles. The DL of the original KL-ONE system [4] contained a constructor called role-value-map that allowed the user to express such relationships. However, it was shown in [14] that role-value-maps make the subsumption problem in KL-ONE undecidable.

The role-value-maps that we consider in the following differ from the ones in [4, 14] in the following respects:

1. Instead of arbitrary role-value-maps of the form $r_1 \circ \cdots \circ r_m \sqsubseteq s_1 \circ \cdots \circ s_n$ we restrict the attention to role-value-maps of the form $r_1 \circ r_2 \sqsubseteq s$, i.e., the right-hand side must be a single role.

2. We consider global role-value-maps, which must hold for all individuals of an interpretation, rather than local ones, which can be asserted selectively for certain individuals.

3. We consider the DL $\mathcal{EL}$, which does not allow value restrictions, whereas the DLs considered in [4, 14] have value restrictions.

The undecidability proof in [14] would also work with the second restriction in place. However, the proof does not work in the presence of the first or the third restriction. Role-value-maps satisfying the first and the second restriction have recently drawn considerable attention [6, 19, 8]. However, for the expressive DLs usually considered there, subsumption easily becomes undecidable [6, 19], and it is quite hard to obtain decidable special cases [8].

For $\mathcal{EL}$ (with or without cyclic terminologies), things are a lot simpler. Not only does subsumption remain decidable, it even stays polynomial when we add role-value-maps satisfying the first two restrictions. We will also show that subsumption becomes undecidable if one adds arbitrary global role-value-maps to $\mathcal{EL}$ (even without cyclic terminologies).

Definition 51 A (global) *role-value-map* is an expression of the form $r_1 \circ \cdots \circ r_m \sqsubseteq s_1 \circ \cdots \circ s_n$ where $m, n \geq 1$ and $r_1, \ldots, s_n$ are role names. It is satisfied in an interpretation $\mathcal{I}$ iff $r_1^{\mathcal{I}} \circ \cdots \circ r_m^{\mathcal{I}} \subseteq s_1^{\mathcal{I}} \circ \cdots \circ s_n^{\mathcal{I}}$, where $\circ$ denotes composition of binary relations. We say that this role-value-map is *restricted* if $m = 2$ and $n = 1$.[4] A finite set of restricted role-value-maps is called an *RBox*. The interpretation $\mathcal{I}$ is a model of the RBox $\mathcal{R}$ iff $\mathcal{I}$ satisfies every role-value-map in $\mathcal{R}$.

Given an $\mathcal{EL}$-TBox $\mathcal{T}$ and an RBox $\mathcal{R}$, subsumption w.r.t. $\mathcal{T}$ and $\mathcal{R}$ is defined in the obvious way:

Definition 52 Let A, B be defined concepts in $\mathcal{T}$. Then

- $A \sqsubseteq_{\mathcal{T}}^{\mathcal{R}} B$ iff $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ holds for all models of $\mathcal{T}$ and $\mathcal{R}$.
- $A \sqsubseteq_{gfp,\mathcal{T}}^{\mathcal{R}} B$ iff $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ holds for all gfp-models of $\mathcal{T}$ that are models of $\mathcal{R}$.

In order to solve the subsumption problem w.r.t. a cyclic $\mathcal{EL}$-TBox $\mathcal{T}$ and an RBox $\mathcal{R}$, we view the restricted role-value-maps $r \circ s \sqsubseteq t \in \mathcal{R}$ as rules that add new edges to $\mathcal{G}_{\mathcal{T}}$.

Definition 53 We say that the role-value-map $r \circ s \sqsubseteq t$ *applies* to the $\mathcal{EL}$-description graph $\mathcal{G}$ iff $\mathcal{G}$ contains edges (u, r, v) and (v, s, w), but does not contain the edge (u, t, w). An *application* of this rule then adds the edge (u, t, w).

Given an $\mathcal{EL}$-description graph $\mathcal{G}$ and an RBox $\mathcal{R}$, we can iterate the application of the role-value-maps in $\mathcal{R}$ to $\mathcal{G}$ until no role-value-map applies. We call the $\mathcal{EL}$-description graph $\widehat{\mathcal{G}}$ obtained this way the *completion* of $\mathcal{G}$ w.r.t. $\mathcal{R}$.

Lemma 54 *Given a finite $\mathcal{EL}$-description graph $\mathcal{G}$ and an RBox $\mathcal{R}$, the completion $\widehat{\mathcal{G}}$ of $\mathcal{G}$ w.r.t. $\mathcal{R}$ always exists, is unique, and can be computed in polynomial time.*

Proof. The applicability of role-value-maps to a graph is monotonic in the following sense: if the role-value-map $r \circ s \sqsubseteq t$ applies to the edges (u, r, v) and (v, s, w) in $\mathcal{G}$, and $\mathcal{G}'$ is obtained from $\mathcal{G}$ by applying some role-value-map, then $r \circ s \sqsubseteq t$ is still applicable to the edges (u, r, v) and (v, s, w) in $\mathcal{G}'$ (since no edges have been removed), unless $\mathcal{G}'$ already contains the edge (u, t, w). Thus, the order of applications of role-value-maps to the graph is irrelevant, which shows uniqueness.

[4]The restriction $m = 2$ is not really necessary. It is easy to see that all our results would still hold if the left-hand sides were compositions of $m \geq 1$ roles. However, the restriction $n = 1$ is vital (see Theorem 60 below).

The application of a role-value-map does not add new nodes to the graph. Thus, if n is the number of nodes in the original graph $\mathcal{G}$ and m is the number of roles occurring in role-value-maps, then at most $n^2 \cdot m$ edges can be added. This implies that an exhaustive application of role-value-maps to the graph $\mathcal{G}$ terminates after at most $n^2 \cdot m$ applications of rules. Consequently, the completion $\widehat{\mathcal{G}}$ exists and can be computed in polynomial time. □

Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{R}$ an RBox, and $\widehat{\mathcal{G}}_\mathcal{T}$ the completion of $\mathcal{G}_\mathcal{T}$ w.r.t. $\mathcal{R}$. The $\mathcal{EL}$-description graph $\widehat{\mathcal{G}}_\mathcal{T}$ corresponds to a TBox $\widehat{\mathcal{T}}$ (i.e., there is a TBox $\widehat{\mathcal{T}}$ such that $\widehat{\mathcal{G}}_\mathcal{T} = \mathcal{G}_{\widehat{\mathcal{T}}}$). We call this TBox the *completion* of $\mathcal{T}$ w.r.t. $\mathcal{R}$.

Lemma 55 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{R}$ an RBox, and $\widehat{\mathcal{T}}$ the completion of $\mathcal{T}$ w.r.t. $\mathcal{R}$. If $\mathcal{I}$ is a model of $\mathcal{R}$, then the following are equivalent:*

- *$\mathcal{I}$ is a model of $\mathcal{T}$.*
- *$\mathcal{I}$ is a model of $\widehat{\mathcal{T}}$.*

Proof. Since $\widehat{\mathcal{G}}_\mathcal{T}$ is obtained from $\mathcal{G}_\mathcal{T}$ by a finite number of applications of role-value-maps, it is enough to show the following: Assume that $\mathcal{G}'$ is obtained from $\mathcal{G}_\mathcal{T}$ by applying the role-value-map $r \circ s \sqsubseteq t \in \mathcal{R}$ to the edges (A, r, B) and (B, s, B') in $\mathcal{G}_\mathcal{T}$, and let $\mathcal{T}'$ be the TBox corresponding to $\mathcal{G}'$. If $\mathcal{I}$ is a model of $\mathcal{R}$, then $\mathcal{I}$ is a model of $\mathcal{T}$ iff it is a model of $\mathcal{T}'$.

The only difference between $\mathcal{T}$ and $\mathcal{T}'$ is that the definition of A in $\mathcal{T}$ (say $A \equiv D$) is extended in $\mathcal{T}'$ by an additional conjunct $\exists t.B'$ (i.e., it is of the form $A \equiv D \sqcap \exists t.B'$).

The existence of the edges (A, r, B) and (B, s, B') in $\mathcal{G}$ implies that $A \sqsubseteq_\mathcal{T} \exists r.\exists s.B'$. Since $\mathcal{R}$ contains the role-value-map $r \circ s \sqsubseteq t$, this implies that $A \sqsubseteq^\mathcal{R}_\mathcal{T} \exists t.B'$. Thus, if $\mathcal{I}$ is a model of $\mathcal{R}$ and $\mathcal{T}$, then is satisfies $A^\mathcal{I} \subseteq (\exists t.B')^\mathcal{I}$. This shows that $A^\mathcal{I} = D^\mathcal{I} = D^\mathcal{I} \cap (\exists t.B')^\mathcal{I} = (D \sqcap \exists t.B')^\mathcal{I}$, and thus $\mathcal{I}$ is also a model of $\mathcal{T}'$.

Conversely, if $\mathcal{I}$ is a model of $\mathcal{R}$ and $\mathcal{T}'$, then $A^\mathcal{I} = D^\mathcal{I} \cap (\exists t.B')^\mathcal{I}$. In addition, the existence of the edges (A, r, B) and (B, s, B') in $\mathcal{G}$ implies that $D^\mathcal{I} \subseteq (\exists r.\exists s.B')^\mathcal{I}$ (since these edges come from D). Since $\mathcal{I}$ is a model of $\mathcal{R}$, this implies $D^\mathcal{I} \subseteq (\exists t.B')^\mathcal{I}$. Consequently, $A^\mathcal{I} = D^\mathcal{I} \cap (\exists t.B')^\mathcal{I} = D^\mathcal{I}$, and thus $\mathcal{I}$ is also a model of $\mathcal{T}$. □

In order to test subsumption w.r.t. $\mathcal{T}$ and $\mathcal{R}$, we compute the completion $\widehat{\mathcal{T}}$ of $\mathcal{T}$ w.r.t. $\mathcal{R}$, and then test subsumption w.r.t. $\widehat{\mathcal{T}}$.

Theorem 56 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{R}$ an RBox, and $\widehat{\mathcal{T}}$ the completion of $\mathcal{T}$ w.r.t. $\mathcal{R}$. Then the following are equivalent for all defined concepts A, B:*

1. $A \sqsubseteq^{\mathcal{R}}_{gfp,\mathcal{T}} B$.
2. $A \sqsubseteq_{gfp,\widehat{\mathcal{T}}} B$.

Proof. ($2 \Rightarrow 1$) Assume that $A \sqsubseteq_{gfp,\widehat{\mathcal{T}}} B$. Let $\mathcal{I}$ be a gfp-model of $\mathcal{T}$ that is a model of $\mathcal{R}$. We must show that $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$. Assume that this gfp-model is based on the primitive interpretation $\mathcal{J}$. Note that the fact that $\mathcal{I}$ is a model of $\mathcal{R}$ depends only on $\mathcal{J}$ (since $\mathcal{J}$ already fixes the interpretation of the roles).

As an easy consequence of Lemma 55 we obtain that $\mathcal{I}$ is also a gfp-model of $\widehat{\mathcal{T}}$. In fact, Lemma 55 shows that $\mathcal{I}$ is a model of $\widehat{\mathcal{T}}$. It remains to be shown that it is the greatest model based on $\mathcal{J}$. Assume that $\mathcal{I}'$ is a model of $\widehat{\mathcal{T}}$ that is based on $\mathcal{J}$, but larger that $\mathcal{I}$. Then $\mathcal{I}'$ is also a model of $\mathcal{R}$ (since this depends only on $\mathcal{J}$). But then Lemma 55 implies that $\mathcal{I}'$ is a model of $\mathcal{T}$, which contradicts our assumption that $\mathcal{I}$ is a gfp-model of $\mathcal{T}$ based on $\mathcal{J}$.

Since $\mathcal{I}$ is a gfp-model of $\widehat{\mathcal{T}}$, $A \sqsubseteq_{gfp,\widehat{\mathcal{T}}} B$ implies $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$.

($1 \Rightarrow 2$) Assume that $A \not\sqsubseteq_{gfp,\widehat{\mathcal{T}}} B$. Then Theorem 13 implies that there is no simulation $Z\colon \mathcal{G}_{\widehat{\mathcal{T}}} \overset{\rightarrow}{\sim} \mathcal{G}_{\widehat{\mathcal{T}}}$ such that $(B, A) \in Z$. We may view $\mathcal{G}_{\widehat{\mathcal{T}}}$ as the graph of a primitive interpretation $\mathcal{J}$, i.e. $\mathcal{G}_{\widehat{\mathcal{T}}} = \mathcal{G}_{\mathcal{J}}$. Let $\mathcal{I}$ be the gfp-model of $\widehat{\mathcal{T}}$ based on $\mathcal{J}$. Then Proposition 15 implies that $A \notin B^{\mathcal{I}}$. Since the identity is a simulation from $\mathcal{G}_{\widehat{\mathcal{T}}}$ to $\mathcal{G}_{\widehat{\mathcal{T}}} = \mathcal{G}_{\mathcal{J}}$ containing the tuple (A, A), we know that $A \in A^{\mathcal{I}}$.

If we can show that $\mathcal{I}$ is a gfp-model of $\mathcal{T}$ that is a model of $\mathcal{R}$, then this implies that $A \not\sqsubseteq^{\mathcal{R}}_{gfp,\mathcal{T}} B$. However, since $\widehat{\mathcal{T}}$ is complete w.r.t. $\mathcal{R}$, $\mathcal{J}$ (and thus $\mathcal{I}$) obviously satisfies all role-value-maps in $\mathcal{R}$. Finally, using Lemma 55 it is easy to show that $\mathcal{I}$ is also a gfp-model of $\mathcal{T}$. □

Since the completion $\widehat{\mathcal{T}}$ of an $\mathcal{EL}$-TBox $\mathcal{T}$ can be computed in polynomial time, and since subsumption w.r.t. gfp-semantics in $\mathcal{EL}$ can be decided in polynomial time, we have the following corollary.

Corollary 57 *The subsumption problem w.r.t. gfp-semantics in $\mathcal{EL}$ remains polynomial in the presence of RBoxes.*

Subsumption w.r.t. descriptive semantics can be treated similarly.

Theorem 58 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{R}$ an RBox, and $\widehat{\mathcal{T}}$ the completion of $\mathcal{T}$ w.r.t. $\mathcal{R}$. Then the following are equivalent for all defined concepts A, B:*

1. $A \sqsubseteq^{\mathcal{R}}_{\mathcal{T}} B$.
2. $A \sqsubseteq_{\widehat{\mathcal{T}}} B$.

Proof. ($2 \Rightarrow 1$) This can be proved just as in the proof of Theorem 56.

($1 \Rightarrow 2$) Assume that $A \sqsubseteq^{\mathcal{R}}_{\mathcal{T}} B$. As in the proof of Theorem 56, we view $\mathcal{G}_{\widehat{\mathcal{T}}}$ as the graph of a primitive interpretation $\mathcal{J}$, i.e. $\mathcal{G}_{\widehat{\mathcal{T}}} = \mathcal{G}_{\mathcal{J}}$. Since $\widehat{\mathcal{T}}$ is complete w.r.t. $\mathcal{R}$, the primitive interpretation $\mathcal{J}$ obviously satisfies all role-value-maps in $\mathcal{R}$.

As done in the proof of ($1 \Rightarrow 2$) of Theorem 29 in [1], we construct a model $\mathcal{I}$ of $\widehat{\mathcal{T}}$ that is based on $\mathcal{J}$. Lemma 34 in [1] shows that $A \in A^{\mathcal{I}}$. Since $\mathcal{J}$ (and thus also $\mathcal{I}$) is a model of $\mathcal{R}$, Lemma 55 above yields that $\mathcal{I}$ is also a model of $\mathcal{T}$. Consequently, $A \sqsubseteq^{\mathcal{R}}_{\mathcal{T}} B$ implies that $A \in B^{\mathcal{I}}$.

We can now proceed as in the proof of Lemma 35 in [1] to show that the simulation $Y\colon \mathcal{G}_{\widehat{\mathcal{T}}} \overset{\rightarrow}{\sim} \mathcal{G}_{\widehat{\mathcal{T}}}$ defined in [1] is a (B, A)-synchronized simulation satisfying $(B, A) \in Y$. By Theorem 19, this implies $A \sqsubseteq_{\widehat{\mathcal{T}}} B$. □

Since the completion $\widehat{\mathcal{T}}$ of an $\mathcal{EL}$-TBox $\mathcal{T}$ can be computed in polynomial time, and since subsumption w.r.t. descriptive semantics in $\mathcal{EL}$ can be decided in polynomial time, we have the following corollary.

Corollary 59 *The subsumption problem w.r.t. descriptive semantics in $\mathcal{EL}$ remains polynomial in the presence of RBoxes.*

The main restriction on the role-value-maps allowed to occur in RBoxes is that the right-hand side must consist of a single role. If we allow for arbitrary role-value-maps, then subsumption becomes undecidable.

Theorem 60 *Subsumption in $\mathcal{EL}$ becomes undecidable in the presence of general (global) role-value-maps.*

Proof. We reduce the word problem for semigroups [11] to the subsumption problem in $\mathcal{EL}$ with general (global) role-value-maps.

Let Σ be a finite alphabet. A semi-Thue system (STS) over Σ is a finite set of rules of the form $x \to y$ where $x, y \in \Sigma^+$. Given an STS T and two words $u, v \in \Sigma^+$ we write $u \to_T v$ iff there is a rule $x \to y \in T$ and words $u_1, u_2 \in \Sigma^*$ such that $u = u_1 x u_2$ and $v = u_1 y u_2$. Let $\sim_T$ denote the reflexive, transitive, and symmetric closure of $\to_T$. The relation $\sim_T$ is an equivalence relation that is compatible with concatenation of words, i.e., $u \sim_T u'$ and $v \sim_T v'$ imply that $uv \sim_T u'v'$. By $[u]_T$ we denote the $\sim_T$-equivalence class of the word u. Concatenation induces a binary associative operation on these classes:

$$[u]_T \cdot [v]_T := [uv]_T.$$

Thus the equivalence classes of words in Σ^+ together with this operation form a semigroup. We call this the *semigroup presented by T*.

The *word problem* for (the semigroup presented by) T is the following question: given words $u, v \in \Sigma^+$, does $u \sim_T v$ hold or not. It is well-known that this problem is in general undecidable [11].

In our reduction, we view the elements of Σ as role names. A non-empty word $w = r_1 \ldots r_m$ over Σ then stands for the composition $r_1 \circ \cdots \circ r_m$ of the roles $r_1, \ldots, r_m$. If $\mathcal{I}$ is an interpretation, the $w^\mathcal{I}$ stands for $r_1^\mathcal{I} \circ \cdots \circ r_m^\mathcal{I}$. Given a word $w = r_1 \ldots r_m$ over Σ, we abbreviate $\exists r_1.\exists r_2.\ldots\exists r_m.C$ by $\exists w.C$.

A given STS T induces the following set of role-value-maps:

$$\mathcal{R}_T := \{x \sqsubseteq y, y \sqsubseteq x \mid x \to y \in T\}.$$

Given two word $u, v \in \Sigma^+$, we define the $\mathcal{EL}$-TBox

$$\mathcal{T}_{u,v} := \{A \equiv \exists u.P,\ B \equiv \exists v.P\}.$$

Since $\mathcal{T}_{u,v}$ is acyclic, descriptive semantics coincides with gfp-semantics.

Claim 1: If A is subsumed by B w.r.t. $\mathcal{T}_{u,v}$ and $\mathcal{R}_T$, then $u \sim_T v$.

Proof of Claim 1. Assume that A is subsumed by B w.r.t. $\mathcal{T}_{u,v}$ and $\mathcal{R}_T$. We us the semigroup $\mathcal{S}$ presented by T to define a model of $\mathcal{T}_{u,v}$ and $\mathcal{R}_T$. Let S be the carrier set of $\mathcal{S}$, i.e., $S = \{[w]_T \mid w \in \Sigma^+\}$.

We define

$$\Delta^\mathcal{I} := \{d_0\} \cup S,$$

and for every role $r \in \Sigma$

$$r^\mathcal{I} := \{(d_0, [r]_T)\} \cup \{([w]_T, [wr]_T) \mid w \in \Sigma^+\}.$$

This definition implies that all roles are interpreted by functional relations. It is easy to show that

$$(*) \qquad w^\mathcal{I} = \{(d_0, [w]_T)\} \cup \{([w']_T, [w']_T \cdot [w]_T) \mid w' \in \Sigma^+\}$$

holds for all words $w \in \Sigma^+$. In addition, we define

$$P^\mathcal{I} = \{[u]_T\}.$$

Finally, by defining

$$A^\mathcal{I} := (\exists u.P)^\mathcal{I} \text{ and } B^\mathcal{I} := (\exists v.P)^\mathcal{I}$$

we make sure that $\mathcal{I}$ is a model of $\mathcal{T}_{u,v}$.

First, we show that $\mathcal{I}$ is also a model of $\mathcal{R}_T$. Given $x \to y \in T$, we must show that $x^\mathcal{I} \subseteq y^\mathcal{I}$ and $y^\mathcal{I} \subseteq x^\mathcal{I}$. By definition of $\sim_T$, $x \to y \in T$ implies that $x \sim_T y$, and thus $[x]_T = [y]_T$. Consequently $x^\mathcal{I} = y^\mathcal{I}$ is an easy consequence of $(*)$ above.

Second, we know that $d_0 \in A^\mathcal{I}$ since $(d_0, [u]_T) \in u^\mathcal{I}$ and $[u]_T \in P^\mathcal{I}$. Since A is subsumed by B w.r.t. $\mathcal{T}_{u,v}$ and $\mathcal{R}_T$, this implies $d_0 \in B^\mathcal{I} = (\exists v.P)^\mathcal{I}$. Since the only element that can be reached from d_0 via $v^\mathcal{I}$ is $[v]_T$, this implies $[v]_T \in P^\mathcal{I} = \{[u]_T\}$, and thus $[u]_T = [v]_T$, i.e., $u \sim_T v$. This completes the proof of Claim 1.

Claim 2: If $u \sim_T v$, then A is subsumed by B w.r.t. $\mathcal{T}_{u,v}$ and $\mathcal{R}_T$.

Proof of Claim 2. Assume that $u \sim_T v$. Then there are a no-negative integer $k \geq 0$ and words $u_0, \ldots, u_k$ such that $u = u_0$, $v = u_k$, and for all $i, 1 \leq i \leq k$, $u_{i-1} \rightarrow_T u_i$ or $u_i \rightarrow_T u_{i-1}$. We prove the claim by induction on k. If $k = 0$, then $u = v$, and the claim is trivially true.

For the induction step, it is sufficient to show the following: if $u \rightarrow_T v$ or $v \rightarrow_T u$, then A is subsumed by B w.r.t. $\mathcal{T}_{u,v}$ and $\mathcal{R}_T$. Since the definition of $\mathcal{R}_T$ is symmetric, it is sufficient to consider the case $u \rightarrow_T v$. Now, $u \rightarrow_T v$ means that there is a rule $x \rightarrow y$ in T such that $u = u_1 x u_2$ and $v = u_1 y u_2$ for some words $u_1, u_2 \in \Sigma^*$.

Assume that $\mathcal{I}$ is a model of $\mathcal{T}_{u,v}$ and $\mathcal{R}_T$, and that $d_0 \in \Delta^\mathcal{I}$ is an element of this model that belongs to $A^\mathcal{I}$. We must show that $d_0 \in B^\mathcal{I}$. Since $A^\mathcal{I} = (\exists u.P)^\mathcal{I}$ and $u = u_1 x u_2$, there are elements $d_1, d_2, d_3 \in \Delta^\mathcal{I}$ such that $(d_0, d_1) \in u_1^\mathcal{I}$, $(d_1, d_2) \in x^\mathcal{I}$, $(d_2, d_3) \in u_2^\mathcal{I}$, and $d_3 \in P^\mathcal{I}$. Since $x \sqsubseteq y \in \mathcal{R}_T$ and $\mathcal{I}$ is a model of $\mathcal{R}_T$, $(d_1, d_2) \in x^\mathcal{I}$ implies $(d_1, d_2) \in y^\mathcal{I}$, and thus $(d_0, d_3) \in v^\mathcal{I}$. This shows that $d_0 \in (\exists v.P)^\mathcal{I} = B^\mathcal{I}$, which completes the proof of Claim 2.

Thus, we have shown that the word problem for semigroups can effectively be reduced to the subsumption problem in $\mathcal{EL}$ with general (global) role-value-maps, which shows that this subsumption problem is undecidable. □

7 Conclusion

Computing the least common subsumer (lcs) and the most specific concept (msc) are important steps in the bottom-up construction of DL knowledge bases. In DLs with existential restrictions, the most specific concept of a given ABox individual need not exist. We have shown that allowing for cyclic definitions with greatest fixpoint (gfp) semantics in the DL $\mathcal{EL}$ overcomes this problem: in this setting, the most specific concept exists and can be computed in polynomial time. But then subsumption and the lcs operation must also be considered w.r.t. cyclic definitions. In [1] it was shown that the subsumption problem remains polynomial if one allows for cyclic definitions in $\mathcal{EL}$. In the present report we have shown that, w.r.t. gfp-semantics, the lcs always exists, and that the binary lcs can be computed in polynomial time.

Subsumption is also polynomial w.r.t. descriptive semantics [1]. For the lcs, descriptive semantics is not that well-behaved: the lcs need not exist in general.

In addition, we could only give a sufficient condition for the existence of the lcs. If this condition applies, then the lcs can be computed in polynomial time. Thus, one of the main technical problems left open by this report is the question how to characterize the cases in which the lcs exists w.r.t. descriptive semantics, and to determine whether in these cases it can always be computed in polynomial time. Another problem that was not addressed by this report is the question of how to characterize and compute the most specific concept w.r.t. descriptive semantics.

We have also shown that adding restricted (global) role-value-maps of the form $r \circ s \sqsubseteq t$ to $\mathcal{EL}$ leaves subsumption polynomial, both w.r.t. descriptive and gfp-semantics. These role-value-maps are of interest in applications in medicine [15]. It should be noted that there are indeed medical application where the expressive power of the small DL $\mathcal{EL}$ appears to be sufficient. In fact, SNOMED, the Systematized Nomenclature of Medicine [5] uses $\mathcal{EL}$ [17, 15, 16].

References

[1] F. Baader. Terminological cycles in a description logic with existential restrictions. LTCS-Report LTCS-02-02, Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology, Germany, 2002. See http://lat.inf.tu-dresden.de/research/reports.html.

[2] Franz Baader and Ralf Küsters. Computing the least common subsumer and the most specific concept in the presence of cyclic $\mathcal{ALN}$-concept descriptions. In *Proc. of the 22nd German Annual Conf. on Artificial Intelligence (KI'98)*, volume 1504 of *Lecture Notes in Computer Science*, pages 129–140. Springer-Verlag, 1998.

[3] Franz Baader, Ralf Küsters, and Ralf Molitor. Computing least common subsumers in description logics with existential restrictions. In *Proc. of the 16th Int. Joint Conf. on Artificial Intelligence (IJCAI'99)*, pages 96–101, 1999.

[4] Ronald J. Brachman and James G. Schmolze. An overview of the KL-ONE knowledge representation system. *Cognitive Science*, 9(2):171–216, 1985.

[5] R.A. Cote, D.J. Rothwell, J.L. Palotay, R.S. Beckett, and L. Brochu. The systematized nomenclature of human and veterinary medicine. Technical report, SNOMED International, Northfield, IL: College of American Pathologists, 1993.

[6] Stephane Demri. The complexity of regularity in grammar logics and related modal logics. *J. of Logic and Computation*, 11(6), 2001.

[7] Monika R. Henzinger, Thomas A. Henzinger, and Peter W. Kopke. Computing simulations on finite and infinite graphs. In *36th Annual Symposium on Foundations of Computer Science*, pages 453–462, Milwaukee, Wisconsin, 1995. IEEE Computer Society Press.

[8] Ian Horrocks and Ulrike Sattler. Description logics with complex role inclusion axioms. LTCS-Report LTCS-02-06, Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology, Germany, 2002. See http://lat.inf.tu-dresden.de/research/reports.html.

[9] Ralf Küsters and Ralf Molitor. Approximating most specific concepts in description logics with existential restrictions. In Franz Baader, Gerd Brewka, and Thomas Eiter, editors, *Proceedings of the Joint German/Austrian Conference on Artificial Intelligence (KI 2001)*, volume 2174 of *Lecture Notes in Artificial Intelligence*, pages 33–47, Vienna, Austria, 2001. Springer-Verlag.

[10] John W. Lloyd. *Foundations of Logic Programming (Second, Extended Edition)*. Springer-Verlag, Berlin, Heidelberg, 1987.

[11] Juri V. Matijasevic. Simple examples of undecidable associative calculi. *Soviet Mathematics (Doklady)*, 8(2):555–557, 1967.

[12] Bernhard Nebel. Terminological cycles: Semantics and computational properties. In John F. Sowa, editor, *Principles of Semantic Networks*, pages 331–361. Morgan Kaufmann, Los Altos, 1991.

[13] Ulrike Sattler. A concept language extended with different kinds of transitive roles. In Günter Görz and Steffen Hölldobler, editors, *Proc. of the 20th German Annual Conf. on Artificial Intelligence (KI'96)*, number 1137 in Lecture Notes in Artificial Intelligence, pages 333–345. Springer-Verlag, 1996.

[14] Manfred Schmidt-Schauß. Subsumption in KL-ONE is undecidable. In Ron J. Brachman, Hector J. Levesque, and Ray Reiter, editors, *Proc. of the 1st Int. Conf. on the Principles of Knowledge Representation and Reasoning (KR'89)*, pages 421–431. Morgan Kaufmann, Los Altos, 1989.

[15] K.A. Spackman. Managing clinical terminology hierarchies using algorithmic calculation of subsumption: Experience with SNOMED-RT. *J. of the American Medical Informatics Association*, 2000. Fall Symposium Special Issue.

[16] K.A. Spackman. Normal forms for description logic expressions of clinical concepts in SNOMED RT. *J. of the American Medical Informatics Association*, pages 627–631, 2001. Symposium Supplement.

[17] K.A. Spackman, K.E. Campbell, and R.A. Cote. SNOMED RT: A reference terminology for health care. *J. of the American Medical Informatics Association*, pages 640–644, 1997. Fall Symposium Supplement.

[18] Alfred Tarski. A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics*, 5:285–309, 1955.

[19] Michael Wessel. Obstacles on the way to qualitative spatial reasoning with description logics: Some undecidability results. In *Proc. of the 2001 Description Logic Workshop (DL 2001)*, 2001.