Restricted 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 paper we have investigated 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.

In this paper 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

In a previous paper [4], we have investigated terminological cycles in the DL $\mathcal{EL}$, which allows for conjunctions, existential restrictions, and the top-concept. In contrast to (even very inexpressive) DLs with value restrictions [1, 2], subsumption in $\mathcal{EL}$ remains polynomial in the presence of terminological cycles for the three types of semantics (least fixpoint (lfp) semantics, greatest fixpoint (gfp) semantics, and descriptive semantics) introduced by Nebel [11].

Although $\mathcal{EL}$ is of a very limited expressive power, there are indeed applications where the small DL $\mathcal{EL}$ appears to be sufficient. In fact, SNOMED, the Systematized Nomenclature of Medicine [6] employs $\mathcal{EL}$ [16, 17]. Even though SNOMED does not appear to use cyclic definitions, this may be due to a lack of technology rather than need. In fact, the Galen medical knowledge base contains many cyclic dependencies [12].

In the medical application mentioned above [15] (but also in other applications [13]) one often 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 this paper 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

*Partially supported by the DFG under grant BA 1122/4-3.

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	$A \equiv D$	$A^{\mathcal{I}} = D^{\mathcal{I}}$

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

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 can 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. The restriction that the right-hand side of role value maps consists of a single role is vital for these results to hold. In fact, we will also show that subsumption in $\mathcal{EL}$ becomes undecidable (even without cyclic terminologies) if general (global) role-value-maps are allowed.

Because of the space constraints, we cannot prove the results in detail. More detailed definitions and proofs can be found in [3].

2 Basic definitions

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 paper, 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 ($\exists r.C$). 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 $\mathcal{T}$.

The semantics of (possibly cyclic) $\mathcal{EL}$-TBoxes we have defined above is called *descriptive semantic* by Nebel [11]. For some applications, it is more appropriate to interpret cyclic concept definitions with the help of appropriate fixpoint semantics. Before we can define these 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

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

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 $\mathcal{I}_1, \mathcal{I}_2 \in Int(\mathcal{J})$, 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 $\preceq_\mathcal{J}$ induces a *complete lattice* on $Int(\mathcal{J})$, i.e., every subset of $Int(\mathcal{J})$ has a least upper bound (lub) and a greatest lower bound (glb). Using *Tarski's fixpoint theorem* [18] for complete lattices, it is not hard to show that, for a given primitive interpretation $\mathcal{J}$, there is always a greatest and a least (w.r.t. $\preceq_\mathcal{J}$) model of $\mathcal{T}$ based on $\mathcal{J}$. We call these models respectively the greatest fixpoint model (gfp-model) and the least fixpoint model (lfp-model) of $\mathcal{T}$. *Greatest (least) fixpoint semantics* considers only gfp-models (lfp-models) as admissible models. In the following, we restrict our attention to gfp- and descriptive semantics since the results in [4] show that cyclic definitions with least fixpoint semantics are not interesting in $\mathcal{EL}$.

Definition 1 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 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 $\mathcal{T}$.

3 Characterizing subsumption in $\mathcal{EL}$

In this section, we recall the characterizations of subsumption w.r.t. gfp-semantics and descriptive semantics developed in [4]. To this purpose, we must represent TBoxes by description graphs, and introduce the notion of a simulation on 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 [4], 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 2 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: $V \to 2^{N_{prim}}$ is a function that labels nodes with sets of primitive concepts.

The normalized 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 $(A, r_1, B_1), \ldots, (A, r_\ell, B_\ell) \in E_\mathcal{T}$.

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

Definition 3 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: $\mathcal{G}_1 \overset{\rightarrow}{\sim} \mathcal{G}_2$ to express that Z is a simulation from $\mathcal{G}_1$ to $\mathcal{G}_2$.

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

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

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

Theorem 4 *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: $\mathcal{G}_\mathcal{T} \stackrel{\rightharpoonup}{\sim} \mathcal{G}_\mathcal{T}$ such that $(B, A) \in Z$.*

This theorem shows that subsumption w.r.t. gfp-semantics in $\mathcal{EL}$ is tractable. In fact, it is easy to see that the set of all simulations from $\mathcal{G}_\mathcal{T}$ to $\mathcal{G}_\mathcal{T}$ is closed under arbitrary unions. Consequently, there always exists a greatest simulation from $\mathcal{G}_\mathcal{T}$ to $\mathcal{G}_\mathcal{T}$. In [8] it is shown that this greatest simulation can be computed in polynomial time. Since there is a simulation Z: $\mathcal{G}_\mathcal{T} \stackrel{\rightharpoonup}{\sim} \mathcal{G}_\mathcal{T}$ such that $(B, A) \in Z$ iff the greatest simulation contains the tuple (B, A), we have the following complexity result for subsumption in $\mathcal{EL}$ w.r.t. gfp-semantics.

Corollary 5 *Subsumption w.r.t. gfp-semantics in $\mathcal{EL}$ can be decided in polynomial time.*

Now, let us turn to subsumption w.r.t. descriptive semantics. 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: $\mathcal{G}_\mathcal{T} \stackrel{\rightharpoonup}{\sim} \mathcal{G}_\mathcal{T}$ 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: $\mathcal{G}_\mathcal{T} \stackrel{\rightharpoonup}{\sim} \mathcal{G}_\mathcal{T}$ a simulation.

Definition 6 The path p_1: $B = B_0 \stackrel{r_1}{\to} B_1 \stackrel{r_2}{\to} B_2 \stackrel{r_3}{\to} B_3 \stackrel{r_4}{\to} \cdots$ in $\mathcal{G}_\mathcal{T}$ is *Z-simulated* by the path p_2: $A = A_0 \stackrel{r_1}{\to} A_1 \stackrel{r_2}{\to} A_2 \stackrel{r_3}{\to} A_3 \stackrel{r_4}{\to} \cdots$ 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 3 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.

Definition 7 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 $(B_n, A_n) \in Z$.

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

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

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 8 Let A, B be defined concepts in $\mathcal{T}$, and $Z\colon \mathcal{G}_\mathcal{T} \stackrel{\rightarrow}{\sim} \mathcal{G}_\mathcal{T}$ 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 recall the characterization of subsumption w.r.t. descriptive semantics from [4].

Theorem 9 *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 $Z\colon \mathcal{G}_\mathcal{T} \stackrel{\rightarrow}{\sim} \mathcal{G}_\mathcal{T}$ such that $(B, A) \in Z$.*

In [4] 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 $Z\colon \mathcal{G}_\mathcal{T} \stackrel{\rightarrow}{\sim} \mathcal{G}_\mathcal{T}$ with $(B, A) \in Z$ can be decided in polynomial time.

Corollary 10 *Subsumption w.r.t. descriptive semantics in $\mathcal{EL}$ can be decided in polynomial time.*

4 Role-value-maps in $\mathcal{EL}$

The DL of the original Kl-One system [5] contained a concept constructor called role-value-map that allowed the user to express relationships between roles. 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 [5, 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, i.e., role-value-maps that 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 [5, 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 [7, 19, 9]. However, for the expressive DLs usually considered there, subsumption easily becomes undecidable [7, 19], and it is quite hard to obtain decidable special cases [9].

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.

Definition 11 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 $n = 1$.[1] 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: 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 12 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 13 *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 time polynomial in the size of $\mathcal{G}$ and $\mathcal{R}$.*

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

[1] 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 for an arbitrary m. However, the restriction $n = 1$ is vital (see Theorem 18 below).

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 15 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{R}$ an RBox, $\widehat{\mathcal{T}}$ the completion of $\mathcal{T}$ w.r.t. $\mathcal{R}$, and A, B defined concepts. Then $A \sqsubseteq^{\mathcal{R}}_{gfp,\mathcal{T}} B$ iff $A \sqsubseteq_{gfp,\widehat{\mathcal{T}}} B$.*

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

Theorem 16 *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox, $\mathcal{R}$ an RBox, $\widehat{\mathcal{T}}$ the completion of $\mathcal{T}$ w.r.t. $\mathcal{R}$, and A, B defined concepts. Then $A \sqsubseteq^{\mathcal{R}}_{\mathcal{T}} B$ iff $A \sqsubseteq_{\widehat{\mathcal{T}}} B$.*

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. gfp- and descriptive semantics in $\mathcal{EL}$ can be decided in polynomial time, we have the following corollary.

Corollary 17 *The subsumption problem w.r.t. gfp-semantics in $\mathcal{EL}$ remains polynomial in the presence of RBoxes. The same is true for the subsumption problem w.r.t. descriptive semantics.*

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

The proof is by reduction of the word problem for semi-Thue systems 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 *word problem* for 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 [10].

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.

Lemma 19 *A is subsumed by B w.r.t. $\mathcal{T}_{u,v}$ and $\mathcal{R}_T$ iff $u \sim_T v$.*

Since the question whether $u \sim_T v$ holds for given words u, v and STS T is in general undecidable, this shows that the subsumption problem becomes undecidable in $\mathcal{EL}$ if one allows for general (global) role-value-maps.

The above reduction is similar to the one used in [14] to show undecidability of subsumption in Kl-One. Note, however, that the result itself is not a consequence of the one in [14] and that there are some differences both in the reduction and in the proof of its correctness due to the fact that we consider a language with existential restrictions, and without value restrictions. It is not clear whether the undecidability result also holds if we consider local role-value-maps in $\mathcal{EL}$. In fact, the technique used in [14] to simulate global role-value-maps with the help of local ones only works in the presence of value restrictions.

References

[1] Franz Baader. Terminological cycles in KL-ONE-based knowledge representation languages. In *Proc. of the 8th Nat. Conf. on Artificial Intelligence (AAAI'90)*, pages 621–626, Boston (Ma, USA), 1990.

[2] Franz Baader. Using automata theory for characterizing the semantics of terminological cycles. *Ann. of Mathematics and Artificial Intelligence*, 18:175–219, 1996.

[3] Franz Baader. Least common subsumers, most specific concepts, and role-value-maps in a description logic with existential restrictions and terminological cycles. LTCS Report 02-07, Theoretical Computer Science, TU Dresden, 2002. Available at http://lat.inf.tu-dresden.de/research/reports.html.

[4] Franz Baader. Terminological cycles in a description logic with existential restrictions. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, 2003. A long version can found as LTCS Report 02-02 at http://lat.inf.tu-dresden.de/research/reports.html.

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

[6] 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.

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

[8] 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.

[9] Ian Horrocks and Ulrike Sattler. Description logics with complex role inclusion axioms. In *Proceedings of the 18th International Joint Conference on Artifi-*

cial Intelligence, 2003. A long version can be found as LTCS Report 02-06 at http://lat.inf.tu-dresden.de/research/reports.html.

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

[11] 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.

[12] A. Rector and I. Horrocks. Experience building a large, re-usable medical ontology using a description logic with transitivity and concept inclusions. In *Proc. of the WS on Ontological Engineering, AAAI Spring Symposium (AAAI'97)*. AAAI Press, 1997.

[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.