RWTH
LTCS–Report

Aachen University of Technology
Research group for
Theoretical Computer Science

Computing Most Specific Concepts in Description Logics with Existential Restrictions

Ralf Küsters and Ralf Molitor

LTCS-Report 00-05

RWTH Aachen
LuFg Theoretische Informatik
http://www-lti.informatik.rwth-aachen.de

Ahornstr. 55
52074 Aachen
Germany

Computing Most Specific Concepts in Description Logics with Existential Restrictions

Ralf Küsters
Institute for Computer Science
and Applied Mathematics
Christian-Albrechts-University of Kiel
kuesters@ti.informatik.uni-kiel.de

Ralf Molitor
Teaching and Reasearch Area for
Theoretical Computer Science
Aachen University of Technology
molitor@cs.rwth-aachen.de

Abstract

Computing the most specific concept (msc) is an inference task that can be used to support the "bottom-up" construction of knowledge bases for KR systems based on description logics. For description logics that allow for number restrictions or existential restrictions, the msc need not exist, though. Previous work on this problem has concentrated on description logics that allow for universal value restrictions and number restrictions, but not for existential restrictions. The main new contribution of this paper is the treatment of description logics with existential restrictions. More precisely, we show that, for the description logic $\mathcal{ALE}$ (which allows for conjunction, universal value restrictions, existential restrictions, negation of atomic concepts, as well as the top and the bottom concept), and its sublanguages $\mathcal{EL}$ (which allows for conjunction, existential restrictions and the top-concept) and $\mathcal{EL}_\neg$ (which extends $\mathcal{EL}$ by negation of atomic concepts) the msc of an ABox-individual only exists in case of acyclic ABoxes. For cyclic ABoxes, we show how to compute an approximation of the msc.

Our approach for computing the msc and its approximation, respectively, is based on representing concept descriptions by certain trees and ABoxes by certain graphs, and then characterizing instance relationships by homomorphisms from trees into graphs. The msc/approximation operation then mainly corresponds to unraveling the graphs into trees and translating them back into concept descriptions.

1 Introduction

Knowledge representation systems based on description logics (DL) can be used to describe the knowledge of an application domain in a structured and formally well-understood way. Traditionally, the knowledge base of a DL system is built in a "top-down" fashion by first formalizing the relevant concepts of the domain (its terminology) by *concept descriptions*, i.e., expressions that are built from atomic concepts (unary predicates) and atomic roles (binary predicates) using the concept constructors provided by the DL language. In a second step, the concept descriptions are used to specify properties of objects and individuals occurring in the domain. DL systems provide their users with inference services that support both steps: classification of concepts and of individuals. Classification of concepts determines subconcept/superconcept relationships (called subsumption relationships) between the concepts of a given terminology, and thus allows one to structure the terminology in the form of a subsumption hierarchy. Classification of individuals (or objects) determines whether a given individual is always an instance of a certain concept (i.e., whether this instance relationship is implied by the descriptions of the individual and the concept).

This traditional "top-down" approach for constructing a DL knowledge base is not always adequate, however. On the one hand, it need not be clear from the outset which are the relevant concepts in a particular application. On the other hand, even if it is clear which (intuitive) concepts should be introduced, it is in general not easy to come up with formal definitions of these concepts within the available description language. For example, in one of our applications in chemical process engineering [13, 4], the process engineers prefer to construct the knowledge base (which consists of descriptions of standard building blocks of process models, such as reactors) in the following "bottom-up" fashion: first, they introduce several "typical" examples of the standard building block as individuals, and then they generalize (the descriptions of) these individuals into a concept description that (i) has all the individuals as instances, and (ii) is the most specific description satisfying property (i). The task of computing a description satisfying (i) and (ii) can be split into two subtasks: computing the most specific concept of a single individual, and computing the least common subsumer of a given finite number of concepts. The *most specific concept* (msc) of an individual b (the *least common subsumer* (lcs) of concept descriptions $C_1, \ldots, C_n$) is the most specific concept description C (expressible in the given description language) that has b as an instance (that subsumes $C_1, \ldots, C_n$).

The second subtask has been thoroughly investigated for several DLs [6, 1, 3, 11, 10]. In this paper, we concentrate on the first subtask. It has been shown [7, 1] that for $\mathcal{ALN}$, the msc of an ABox individual need not exist. In [7], the authors circumvent the problem by using certain approximations of the msc. More precisely, instead of the msc of an ABox individual a, they consider the most specific concept C with depth less or equal to a given bound k such that a is an instance of C. In [1], the authors solve the problem by extending the language to *cyclic* $\mathcal{ALN}$-concept descriptions with greatest fixpoint semantics (see [9] for an automata theoretic characterization of different semantics of cyclic $\mathcal{ALN}$-

concept descriptions). In order to be able to employ the resulting algorithm in the bottom-up construction of knowledge bases, they also give a characterization of the lcs of cyclic concept descriptions.

Since for $\mathcal{ALE}$ cyclic concept descriptions and their semantics are not yet well-investigated, it is an open problem whether this approach can be adapted to $\mathcal{ALE}$. So in this work, we concentrate on the first pragmatic approach. More precisely, we consider the approximation of the msc by concept descriptions of limited depth for the DL employed in our process engineering application. This language, called $\mathcal{ALE}$, allows both for value restrictions and existential restrictions, but not for full negation and disjunction. For instance, it can be used to introduce the concept of a reactor with cooling jacket by the description $\mathsf{Reactor} \sqcap \exists \mathsf{connected\text{-}to.Cooling\text{-}Jacket} \sqcap \forall \mathsf{functionality}.\neg \mathsf{Vaporize}$, where $\mathsf{Vaporize}$ is a primitive concept (i.e., not further defined). Previous work on how to approximate or compute, respectively, the msc [7, 1] has concentrated on sublanguages of the DL used by the system CLASSIC [5], which allows (among other constructors) for value restrictions, but not for existential restrictions. Thus, the main new contribution of the present paper is the treatment of existential restrictions.

We will first show that the msc of an ABox individual a need not exist for the DL $\mathcal{ALE}$ nor the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$. ($\mathcal{EL}$ only allows for conjunction and existential restrictions, and $\mathcal{EL}_\neg$ extends $\mathcal{EL}$ by primitive negation.) For didactic reasons, we will then start with showing how to compute the approximation of the msc in the small language $\mathcal{EL}$, and then extend our treatment in two steps to $\mathcal{EL}_\neg$ and to $\mathcal{ALE}$. For all three languages, the basic idea is as follows: first, introduce a sound and complete characterization of instance relationships that is appropriate in the sense that it allows for characterizing the (approximation of the) msc in a rather intuitive way. In [3] the very same idea is employed for the lcs in $\mathcal{ALE}$: the authors first introduce a representation of concept descriptions by so-called description trees, and then characterize subsumption by the existence of certain homomorphisms between description trees. From this characterization, they deduce that the lcs operation on concept descriptions corresponds to the product operation on description trees, which can easily be computed.

Adapting the characterization of subsumption to instance relationships leads to representing ABoxes by so-called description graphs and characterizing instance by the existence of certain homomorphisms from description trees into description graphs. The characterization of instance obtained this way is sound and complete only for $\mathcal{EL}$, but incomplete for $\mathcal{EL}_\neg$ and $\mathcal{ALE}$. For $\mathcal{EL}$, we will deduce that computing the approximation of the msc of a corresponds to (i) unraveling the description graph into a tree with root a, (ii) truncating the resulting tree at the given depth k, and (iii) translating the $\mathcal{EL}$-description tree obtained this way into an $\mathcal{EL}$-concept description. In case of an acyclic ABox, the tree obtained in step (i) is finite and even yields the msc of a. For $\mathcal{EL}_\neg$, the characterization of instance relationships requires the existence of a certain homomorphism not only into the description graph obtained from the ABox, but into each so-called primitive completion of this graph. Then, the approximation

Construct name	Syntax	Semantics
top-concept	$\top$	Δ
conjunction	$C \sqcap D$	$C^I \cap D^I$
existential restrictions	$\exists r.C$	$\{x \in \Delta \mid \exists y : (x,y) \in r^I \wedge y \in C^I\}$
value restrictions	$\forall r.C$	$\{x \in \Delta \mid \forall y : (x,y) \in r^I \rightarrow y \in C^I\}$
primitive negation	$\neg P$	$\Delta \setminus P^I$
bottom-concept	$\bot$	$\emptyset$
number restrictions	$(\geq n\ r)$	$\{x \in \Delta \mid \#\{y \mid (x,y) \in r^{\mathcal{I}}\} \geq n\}$
number restrictions	$(\leq n\ r)$	$\{x \in \Delta \mid \#\{y \mid (x,y) \in r^{\mathcal{I}}\} \leq n\}$

Table 1: Syntax and semantics of concept descriptions. P denotes a concept name from N_C, r a role name from N_R, and n a nonnegative integer.

of the msc can be obtained by the steps (i)–(iii) starting with the product of all these primitive completions (which again is a description graph). For acyclic $\mathcal{EL}_\neg$-ABoxes, this again yields the msc. Unfortunately, we could not yet extend this characterization to $\mathcal{ALE}$. However, based on the sound characterization of instance relationships, we will introduce an algorithm for computing an "approximation of the approximation of the msc". Furthermore, we will give some examples that illustrate the main problems one has to solve when searching for a sound and complete characterization of instance in $\mathcal{ALE}$ and an algorithm for computing the (approximations of the) msc in $\mathcal{ALE}$, respectively.

2 Preliminaries and known results

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 work, we consider concept descriptions built from the constructors shown in Table 1. In the description logic $\mathcal{EL}$, concept descriptions are formed using the constructors top-concept ($\top$), conjunction ($C \sqcap D$) and existential restriction ($\exists r.C$). The description logic $\mathcal{EL}_\neg$ additionally provides us with primitive negation ($\neg P$, $P \in N_C$), and $\mathcal{ALE}$ allows for all the constructors shown in Table 1 except of number restrictions. Finally, $\mathcal{ALN}$ allows for the top- and bottom-concept, concept conjunction, primitive negation, value restrictions, and number restrictions.

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

Definition 1 (Subsumption, lcs) *Let C, D, E be concept descriptions of the same DL $\mathcal{L}$.*

1. D subsumes C *(for short $C \sqsubseteq D$) iff $C^I \subseteq D^I$ for all interpretations $\mathcal{I}$.* C *is* equivalent *to D (for short $C \equiv D$) iff $C \sqsubseteq D$ and $D \sqsubseteq C$.*
2. *The concept description E is called* least common subsumer *of C and D iff $C \sqsubseteq E$ and $D \sqsubseteq E$ and for all E' with $C \sqsubseteq E'$ and $D \sqsubseteq E'$, it is $E \sqsubseteq E'$.*

For $\mathcal{ALE}$, subsumption can be characterized by means of *homomorphisms* between *$\mathcal{ALE}$-description trees* [3].

Definition 2 ($\mathcal{ALE}$-description trees) *An* $\mathcal{ALE}$-description tree *is a tree of the form $\mathcal{G} = (V, E, v_0, \ell)$ with root v_0 where*

- *the edges in E are labeled with role names $r \in N_R$ or $\forall r$ for some $r \in N_R$, and*
- *the nodes $v \in V$ are labeled with sets $\ell(v) = \{P_1, \ldots, P_n\}$ where each P_i, $1 \leq i \leq n$, is of one of the following forms: $P_i \in N_C$, $P_i = \neg P$ for some $P \in N_C$, or $P_i = \bot$.*

The empty label corresponds to the top-concept.

Every $\mathcal{ALE}$-concept description can be written (modulo equivalence) as

$$C = Q_1 \sqcap \ldots \sqcap Q_n \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_m.C_m \sqcap \forall s_1.D_1 \sqcap \ldots \sqcap \forall s_k.D_k$$

with $Q_i \in N_C \cup \{\neg P \mid P \in N_C\} \cup \{\bot, \top\}$. The depth $\mathsf{depth}(C)$ of C is defined as the maximal depth of nested quantification in C. Now, the $\mathcal{ALE}$-description tree $\mathcal{G}(C) := (V, E, v_0, \ell)$ corresponding to the $\mathcal{ALE}$-concept description C is inductively defined as follows:

- If $\mathsf{depth}(C) = 0$, then $V := \{v_0\}$, $E := \emptyset$, and $\ell(v_0) := \{Q_1, \ldots, Q_n\} \setminus \{\top\}$;
- otherwise, let $\mathcal{G}_i = (V_i, E_i, v_{0i}, \ell_i)$ be the recursively defined description trees corresponding to C_i, $1 \leq i \leq m$, and $\mathcal{H}_j = (W_j, F_j, w_{0j}, \kappa_j)$ the recursively defined description trees corresponding to D_j, $1 \leq j \leq k$ such that the sets V_i, W_j and $\{v_0\}$ are pairwise disjoint; then
  - $V := \{v_0\} \bigcup_{1 \leq i \leq m} V_i \cup \bigcup_{1 \leq j \leq k} W_j$,
  - $E \;:=\; \{v_0 r_i v_{0i} \mid 1 \leq i \leq m\} \cup \{v_0 \forall s_j w_{0j} \mid 1 \leq j \leq k\} \cup \bigcup_{1 \leq i \leq m} E_i \cup \bigcup_{1 \leq j \leq k} F_j$, and
  - $\ell(v) := \begin{cases} \{Q_1, \ldots, Q_n\} \setminus \{\top\}, & v = v_0 \\ \ell_i(v), & v \in V_i, 1 \leq i \leq m \\ \kappa_j(v), & v \in W_j, 1 \leq j \leq k. \end{cases}$

The depth of a description tree $\mathcal{G}$ is defined as the length of the longest path in $\mathcal{G}$. Now, each $\mathcal{ALE}$-description tree $\mathcal{G} = (V, E, v_0, \ell)$ is inductively translated into an $\mathcal{ALE}$-concept description $C_\mathcal{G}$ as follows:

- If $\mathsf{depth}(\mathcal{G}) = 0$, then $V = \{v_0\}$ and $E = \emptyset$. Define

$$C_\mathcal{G} := \begin{cases} Q_1 \sqcap \ldots \sqcap Q_n, & \ell(v_0) = \{Q_1, \ldots, Q_n\} \\ \top, & \ell(v_0) = \emptyset; \end{cases}$$

- otherwise, let $\ell(v_0) = \{Q_1, \ldots, Q_n\}$, $n \geq 0$, and let $\{v_1, \ldots, v_m\}$ be the set of all successors of v_0 with $v_0 r_i v_i \in E$ and $\{w_1, \ldots, w_k\}$ the set of all successors of v_0 with $v_0 \forall s_j w_j \in E$. Further, let C_i (D_j) denote the recursively defined concept descriptions obtained from the subtrees with root v_i, $1 \leq i \leq m$ (w_j, $1 \leq j \leq k$). Define

$$C_\mathcal{G} := Q_1 \sqcap \ldots \sqcap Q_n \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_m.C_m \sqcap \forall s_1.D_1 \sqcap \ldots \sqcap \forall s_k.D_k.$$

Definition 3 (Homomorphisms between $\mathcal{ALE}$-description trees) *A homomorphism from an $\mathcal{ALE}$-description tree $\mathcal{H} = (V_H, E_H, w_0, \ell_H)$ to an $\mathcal{ALE}$-description tree $\mathcal{G} = (V_G, E_G, v_0, \ell_G)$ is a mapping $\varphi : V_H \longrightarrow V_G$ such that*

1. *$\varphi(w_0) = v_0$,*
2. *for all $v \in V_H$ we have $\ell_H(v) \subseteq \ell_G(\varphi(v))$ or $\ell_G(\varphi(v)) = \{\bot\}$,*
3. *for all $vrw \in E_H$, either $\varphi(v)r\varphi(w) \in E_G$, or $\varphi(v) = \varphi(w)$ and $\ell_G(\varphi(v)) = \{\bot\}$, and*
4. *for all $v\forall rw \in E_H$, either $\varphi(v)\forall r\varphi(w) \in E_G$, or $\varphi(v) = \varphi(w)$ and $\ell_G(\varphi(v)) = \{\bot\}$.*

In order to obtain a sound and complete characterization of subsumption, the concept descriptions must be transformed into certain normal forms before translating them into description trees [3]. The *$\mathcal{ALE}$-normal form* of an $\mathcal{ALE}$-concept description C is obtained from C by exhaustively applying the following normalization rules (modulo commutativity and associativity of conjunction):

$$\begin{array}{lrcl} 1. & \forall r.\top & \longrightarrow & \top \\ 2. & E \sqcap \top & \longrightarrow & E \\ 3. & \forall r.E \sqcap \forall r.F & \longrightarrow & \forall r.(E \sqcap F) \\ 4. & \forall r.E \sqcap \exists r.F & \longrightarrow & \forall r.E \sqcap \exists r.(E \sqcap F) \\ 5. & P \sqcap \neg P & \longrightarrow & \bot, \text{ for each } P \in N_C \\ 6. & \exists r.\bot & \longrightarrow & \bot \\ 7. & E \sqcap \bot & \longrightarrow & \bot. \end{array}$$

The $\top$-*normal form* of C is obtained from C by exhaustively applying th normalization rules 1. and 2. C is said to be in $\mathcal{ALE}$-normal form ($\top$-normal form) if none of the rules (neither rule 1. nor 2.) is applicable to C. Note that the

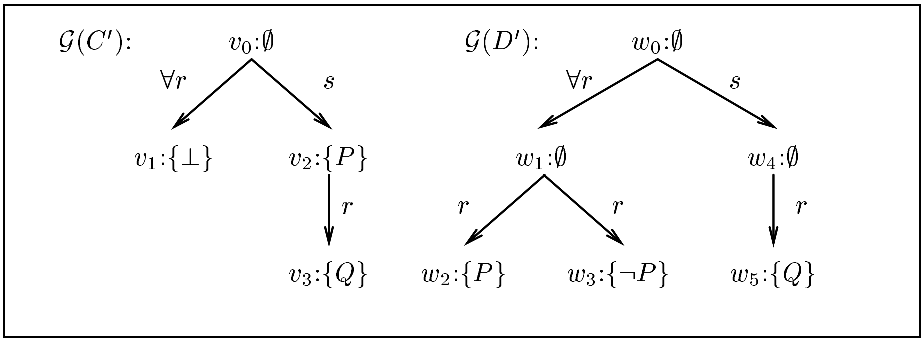

Figure 1: An example for the characterization of subsumption in $\mathcal{ALE}$ by homomorphisms and description trees.

$\mathcal{ALE}$-normal form C' obtained from an $\mathcal{ALE}$-concept description C can be of exponential size w.r.t. the size of C (see [3] for an example). It is easy to see that the $\mathcal{ALE}$-/$\top$-normal form of C is equivalent to C.

Now, subsumption $C \sqsubseteq D$ can be characterized as follows:

Theorem 4 *[3] Let C, D be $\mathcal{ALE}$-concept descriptions, C' the $\mathcal{ALE}$-normal form of C, D' the $\top$-normal form of D, and $\mathcal{G}(C'), \mathcal{G}(D')$ the corresponding $\mathcal{ALE}$-description trees. Then $C \sqsubseteq D$ iff there exists a homomorphism from $\mathcal{G}(D')$ to $\mathcal{G}(C')$.*

Example 5 *Consider the $\mathcal{ALE}$ concept descriptions*

$$\begin{aligned} C &:= \forall r.\exists r.(P \sqcap \neg P) \sqcap \exists s.(P \sqcap \exists r.Q), \\ D &:= \forall r.(\exists r.P \sqcap \exists r.\neg P) \sqcap \exists s.(\forall r.\top \sqcap \exists r.Q). \end{aligned}$$

The $\top$-normal form of D is given by $D' = \forall r.(\exists r.P \sqcap \exists r.\neg P) \sqcap \exists s. \sqcap \exists r.Q$, and the $\mathcal{ALE}$-normal form of C is $C' := \forall r.\bot \sqcap \exists s.(P \sqcap \exists r.Q)$. The corresponding $\mathcal{ALE}$-description trees $\mathcal{G}(C')$ and $\mathcal{G}(D')$ are depicted in Figure 1.

If we define the mapping φ such that it maps w_0 onto v_0; w_1, w_2, and w_3 onto v_1; w_4 onto v_2; and w_5 onto v_3, then it is easy to see that φ yields a homomorphism from $\mathcal{G}(D')$ into $\mathcal{G}(C')$. Thus, Theorem 4 implies $C \sqsubseteq D$.

The inference problem of computing the lcs of $n \geq 2$ concept descriptions has thoroughly been investigated for the DLs $\mathcal{ALE}$ [3], $\mathcal{ALN}$ [6, 1], and Classic [11, 10]. As shown in [3], the lcs of $n \geq 2$ $\mathcal{ALE}$-concept descriptions always exists, and it can be computed in exponential time. The lcs of $n \geq 2$ $\mathcal{ALN}$-concept descriptions also always exists, and it can be computed in polynomial time [6]. Even in the presence of cyclic $\mathcal{ALN}$-concept descriptions, the lcs always exists, and it can be computed in double-exponential time [1]. Things become less rosy, however, if we consider the most specific concept of ABox individuals.

Definition 6 (ABox) *An* ABox $\mathcal{A}$ *is a finite set of assertions of the form* $(a, b) : r$ *(*role assertions*) or* $a : C$ *(*concept assertions*), where* a, b *are* individuals *from a set* N_I*,* r *is a role name, and* C *is a concept description. An ABox is called* $\mathcal{L}$*-ABox if all concept descriptions occurring in* $\mathcal{A}$ *are* $\mathcal{L}$*-concept descriptions.*

In the presence of an ABox, an interpretation $\mathcal{I}$ additionally assigns an element $a^{\mathcal{I}} \in \Delta$ to each individual a occurring in $\mathcal{A}$ such that $A \neq B$ implies $A^{\mathcal{I}} \neq b^{\mathcal{I}}$ (unique name assumption). It is a *model* of $\mathcal{A}$ iff it satisfies $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ for all role assertions $(a, b) : r \in \mathcal{A}$, and $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all concept assertions $a : C \in \mathcal{A}$.

Definition 7 (Instance, msc) *Let* $\mathcal{A}$ *be an* $\mathcal{L}$*-ABox,* a *an individual in* $\mathcal{A}$*, and* C *an* $\mathcal{L}$*-concept description.*

1. *a is an* instance *of C w.r.t. $\mathcal{A}$ ($a \in_{\mathcal{A}} C$) iff $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{A}$.*
2. *C is the* most specific concept *for a w.r.t. $\mathcal{A}$ iff $a \in_{\mathcal{A}} C$ and for all C' with $a \in_{\mathcal{A}} C'$, it is $C \sqsubseteq C'$.*

Depending on the expressive power of the underlying DL $\mathcal{L}$, the msc of an individual a w.r.t. an $\mathcal{L}$-ABox $\mathcal{A}$ need not exist in general. Due to cyclic dependencies between individuals, i.e., cycles build by role assertions in the ABox, it might be the case that there exist infinite many $\mathcal{L}$-concept descriptions a is an instance of, but none most specific concept description with this property. The following example illustrates this situation for $\mathcal{ALN}$ and $\mathcal{ALE}$.

Example 8 *First, consider the $\mathcal{ALN}$-ABox $\mathcal{A} = \{a : P, a : (\leq 1\ r), (a, a) : r\}$. In [1] it is shown that there does not exist the msc of a w.r.t. $\mathcal{A}$: It is easy to see that, for each $n \geq 0$, a is an instance of the $\mathcal{ALN}$-concept description*

$$C_n \ := \ \underbrace{\forall r. \cdots \forall r}_{n\ times}.(P \sqcap (\leq 1\ r) \sqcap (\geq 1\ r)).$$

Intuitively, the msc is thus given as the infinite conjunction $\bigsqcap\limits_{n \geq 0} C_n$. Obviously, this conjunction cannot be represented by an $\mathcal{ALN}$-concept description.

For $\mathcal{ALE}$ and its sublanguages $\mathcal{EL}$ and $\mathcal{EL}_{\neg}$, we encounter the same problem for the even smaller ABox $\mathcal{A}' = \{a : P, (a, a) : r\}$: It is easy to see that, for each $n \geq 0$, a is an instance of the $\mathcal{ALE}$/-$\mathcal{EL}$-/$\mathcal{EL}_{\neg}$-concept description

$$C_n \ := \ \underbrace{\exists r. \cdots \exists r}_{n\ times}.P.$$

Assume that there exists an $\mathcal{ALE}$-/$\mathcal{EL}$-/$\mathcal{EL}_{\neg}$-concept description $C \equiv \mathsf{msc}_{\mathcal{A}'}(a)$. Let $\mathsf{depth}(C) = k$, C' the $\mathcal{ALE}$-normal form of C, and $\mathcal{G}(C')$ the corresponding $\mathcal{ALE}$-description tree. Obviously, there does not exist a homomorphism from $\mathcal{G}(C_{k+1})$ into $\mathcal{G}(C')$. Since C_{k+1} is in $\top$-normal form, Theorem 4 implies $C \not\sqsubseteq C_{k+1}$ in contradiction to $a \in_{\mathcal{A}'} C_{k+1}$ and $C \equiv \mathsf{msc}_{\mathcal{A}'}(a)$.

For cyclic $\mathcal{ALN}$-ABoxes, the msc can be characterized by *cyclic* $\mathcal{ALN}$-concept descriptions. Such concepts are defined by means of a cyclic TBox, and they are interpreted using the greatest fixed-point semantics [1, 9]. In the above example, the concept C defined by the cyclic $\mathcal{ALN}$-TBox

$$\mathcal{T} = \{C \doteq P \sqcap (\leq 1\ r) \sqcap (\geq 1\ r) \sqcap \forall r.C\}$$

yields the msc of a [1].

As already mentioned in the introduction, cyclic concept descriptions are not yet well-investigated for DLs with existential restrictions. Thus, in this work, we concentrate on approximations of the msc in $\mathcal{ALE}$ (and the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$). The approximation of the msc by concept descriptions with limited depth as introduced in [7] is formally defined as follows:

Definition 9 *Let $\mathcal{A}$ be an $\mathcal{L}$-ABox, a an individual in $\mathcal{A}$, C an $\mathcal{L}$-concept description, and $k \in \mathbb{N}$ a nonnegative integer. C is called* k-approximation *of a w.r.t. $\mathcal{A}$ ($C = \mathsf{msc}_{k,\mathcal{A}}(a)$) iff*

1. *$a \in_\mathcal{A} C$,*
2. *$\mathsf{depth}(C) \leq k$, and*
3. *for all C' with $a \in_\mathcal{A} C'$ and $\mathsf{depth}(C') \leq k$, it is $C \sqsubseteq C'$.*

For the ABox $\mathcal{A}'$ from Example 8, the $\mathcal{ALE}$-concept description C_k yields the k-approximation of a w.r.t. $\mathcal{A}$ for each $k \geq 0$.

In the following sections, we will show that, for the DLs $\mathcal{EL}$, $\mathcal{EL}_\neg$, and $\mathcal{ALE}$, the k-approximation of an individual a w.r.t. $\mathcal{A}$ always exists. For $\mathcal{ALE}$, however, we only have a very inefficient algorithm, since the characterization of instance relationships underlying the more efficient algorithms introduced for $\mathcal{EL}$ and $\mathcal{EL}_\neg$ could not be adapted to $\mathcal{ALE}$.

3 Most Specific Concepts in $\mathcal{EL}$

First, we introduce the characterization of instance in $\mathcal{EL}$ that will be used to prove soundness and completeness of the approximation algorithm presented in Section 3.2.

3.1 Characterizing instance in $\mathcal{EL}$

The characterization of instance can be seen as an extension of the characterization of subsumption given in Theorem 4 (see also[3]). Roughly speaking, the idea is to translate the ABox $\mathcal{A}$ into a so-called $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$, and then to characterize $a \in_\mathcal{A} C$ by the existence of a homomorphism φ from the $\mathcal{EL}$-description tree $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ such that the root of $\mathcal{G}(C)$ is mapped onto a.

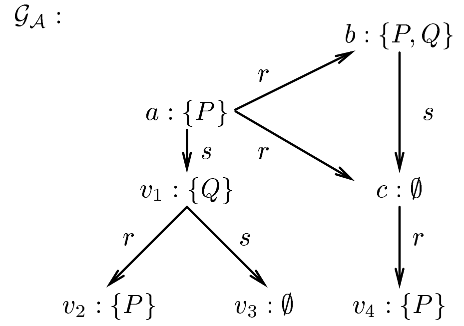

Figure 2: The $\mathcal{EL}$-description graph corresponding to $\mathcal{A}$.

Definition 10 ($\mathcal{EL}$-description tree/$\mathcal{EL}$-description graph) *An* $\mathcal{EL}$-description tree *(*$\mathcal{EL}$-description graph*) is a tree (graph) of the form $\mathcal{G} = (V, E, v_0, \ell)$ with root v_0 ($\mathcal{G}(V, E, \ell)$) where*

- *the edges in E are labeled with role names $r \in N_R$, and*
- *the nodes $v \in V$ are labeled with subsets of N_C, i.e., $\ell(v) \subseteq N_C$ for all $v \in V$.*

The empty label corresponds to the top-concept.

Note that $\mathcal{EL}$-concept descriptions trivially satisfy the conditions on the $\top$- and $\mathcal{ALE}$-normal form. Thus, Theorem 4 yields

Corollary 11 *Let C, D be $\mathcal{EL}$-concept descriptions and $\mathcal{G}(C), \mathcal{G}(D)$ the corresponding $\mathcal{EL}$-description trees. Then, $C \sqsubseteq D$ iff there exists a homomorphism φ from $\mathcal{G}(D)$ into $\mathcal{G}(C)$.*

The graphical representation of an $\mathcal{EL}$-ABox $\mathcal{A}$ yields the starting point for the definition of the $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ corresponding to $\mathcal{A}$, i.e., the individuals in $\mathcal{A}$ yield a subset of the nodes in $\mathcal{G}(\mathcal{A})$ and the role assertions in $\mathcal{A}$ yield some of the edges in $\mathcal{G}(\mathcal{A})$. Concept assertions $a : C \in \mathcal{A}$ are translated as follows: the concept names occurring on the top-level of C yield the label of a, and for each existential restriction $\exists r.C'$ on the top-level of C, the $\mathcal{EL}$-description tree $\mathcal{G}(C') = (V_{C'}, E_{C'}, v_{0C'}, \ell_{C'})$ is added to the graph together with the edge $arv_{0C'}$. For example, the $\mathcal{EL}$-ABox

$$\mathcal{A} \quad = \quad \{a : P \sqcap \exists s.(Q \sqcap \exists r.P \sqcap \exists s.\top), b : P \sqcap Q, c : \exists r.P, (a, b) : r, (a, c) : r, (b, c) : s\}$$

yields the $\mathcal{EL}$-description graph depicted in Figure 2. Formally, $\mathcal{G}(\mathcal{A})$ is defined as follows:

Definition 12 ($\mathcal{EL}$-description graph of an $\mathcal{EL}$-ABox) *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox and $\mathsf{Ind}(\mathcal{A})$ the set of individuals occurring in $\mathcal{A}$. For each $a \in \mathsf{Ind}(\mathcal{A})$, let $C_a := \bigsqcap_{a:D\in\mathcal{A}} D$, if there exists a concept assertion $a : D \in \mathcal{A}$; otherwise, $C_a := \top$. Finally, $\mathcal{G}(C_a) = (V_a, E_a, a, \ell_a)$ denotes the $\mathcal{EL}$-description trees corresponding to C_a, $a \in \mathsf{Ind}(\mathcal{A})$, where w.l.o.g. the sets V_a are assumed to be pairwise disjoint.*

The $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ of $\mathcal{A}$ *is defined by $\mathcal{G}(\mathcal{A}) := (V, E, \ell)$ with*

- $V := \bigcup_{a\in\mathsf{Ind}(\mathcal{A})} V_a$,
- $E := \{arb \mid (a, b) : r \in \mathcal{A}\} \cup \bigcup_{a\in\mathsf{Ind}(\mathcal{A})} E_a$, *and*
- $\ell(v) := \ell_a(v)$ *for all* $v \in V_a$.

It remains to adapt the notion of a homomorphism to $\mathcal{EL}$-description graphs and trees: A mapping $\varphi : V_C \longrightarrow V$ is a *homomorphism* from the $\mathcal{EL}$-description tree $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ into the $\mathcal{EL}$-description graph $\mathcal{G} = (V, E, \ell)$ if

1. $\ell_C(v) \subseteq \ell(\varphi(v))$ for all $v \in V_C$, and
2. $\varphi(v)r\varphi(w) \in E$ for all $vrw \in E_C$.

The first condition of Definition 3 is now directly integrated into the characterization of instance relationships in $\mathcal{EL}$.

Theorem 13 *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$ an individual in $\mathcal{A}$, and C an $\mathcal{EL}$-concept description. Further, let $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$ denote the $\mathcal{EL}$-description graph of $\mathcal{A}$ and $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ the $\mathcal{EL}$-description tree corresponding to C. Then, $a \in_{\mathcal{A}} C$ iff there exists a homomorphism $\varphi : V_C \longrightarrow V$ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$.*

Proof of the if-direction: As already mentioned in the introduction, this characterization of instance is also sound for $\mathcal{ALE}$ (using appropriate definitions of $\mathcal{ALE}$-description graphs and homomorphisms from $\mathcal{ALE}$-description trees into $\mathcal{ALE}$-description graphs). This soundness result will be formalized in Lemma 36 in Section 5. The if-direction of Theorem 13 is an immediate consequence of this lemma.
Proof of the only-if-direction: For this poof, we need the canonical interpretation induced by an $\mathcal{EL}$-description graph.

Definition 14 (Canonical interpretation) *Let $\mathcal{G} = (V, E, \ell)$ be an $\mathcal{EL}$-description graph. The* canonical interpretation $\mathcal{I}(\mathcal{G})$ *is defined by $\mathcal{I}(\mathcal{G}) := (\Delta_{\mathcal{I}(\mathcal{G})}, \cdot^{\mathcal{I}(\mathcal{G})})$ with*

- $\Delta_{\mathcal{I}(\mathcal{G})} := V$,
- $P^{\mathcal{I}(\mathcal{G})} := \{v \in V \mid P \in \ell(v)\}$ *for all* $P \in N_C$, *and*
- $r^{\mathcal{I}(\mathcal{G})} := \{(v, w) \in V \times V \mid (vrw \in E)\}$ *for all* $r \in N_R$.

The following lemma formalizes the important property used in the proof of completeness, namely that, for an $\mathcal{EL}$-ABox, the canonical interpretation $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ is a model of $\mathcal{A}$.

Lemma 15 *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox and $\mathcal{G}(\mathcal{A})$ the corresponding $\mathcal{EL}$-description graph. The canonical interpretation $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ is a model of $\mathcal{A}$.*

Proof: We have to show that $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ satisfies each assertion in $\mathcal{A}$.

By construction, $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ satisfies each role assertion in $\mathcal{A}$.

Let $a : D \in \mathcal{A}$ where $D = P_1 \sqcap \ldots \sqcap P_n \sqcap \exists r_1.D_1 \sqcap \ldots \sqcap \exists r_m.D_m$, $P_i \in N_C$. We show that

1. $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in P_i^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq i \leq n$, and
2. $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in (\exists r_j.D_j)^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq j \leq m$.

Ad (1): By definition of $\mathcal{G}(\mathcal{A})$, it is $P_i \in \ell(a)$, and by definition of $\mathcal{I}(\mathcal{G}(\mathcal{A}))$, we get $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in P_i^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq i \leq n$.

Ad (2): Let C_a be defined as in Definition 12. By definition of $\mathcal{G}(\mathcal{A})$, for each $1 \leq j \leq m$, there exists a node $v_j \in V$ such that $ar_jv_j \in E$ and $D_j \equiv C_{\mathcal{G}(C_a)(v_j)}$, where $\mathcal{G}(C_a)(v_j)$ denotes the subtree with root v_j of $\mathcal{G}(C_a)$. By induction on the depth of D_j, it is easy to see that $v_j^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in (C_{\mathcal{G}(C_a)(v_j)})^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ (see, for example, the proof of Lemma 37 in [2]). In addition, it is $(a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}, v_j^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}) \in r_j^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$, and thus $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in (\exists r_j.D_j)^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$.

Now, (1) and (2) imply $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in D^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$. □

In order to complete the proof of the only-if-direction, we will show the following

Claim: Let $v \in V$. If $v \in D^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$, then there exists a homomorphism φ from $\mathcal{G}(D) = (V_D, E_D, w_0, \ell_D)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = v$.

Since $a \in_{\mathcal{A}} C$ and $\mathcal{I}(\mathcal{G}(\mathcal{A})) \models \mathcal{A}$, the claim implies that there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$.

We prove the claim by induction on $\mathsf{depth}(D)$.

$\mathsf{depth}(D) = 0$, i.e., $D = P_1 \sqcap \ldots \sqcap P_n$. Then $\mathcal{G}(D) = (\{w_0\}, \emptyset, w_0, \ell_D)$ with $\ell_D(w_0) = \{P_1, \ldots, P_n\}$. Define φ by $\varphi(w_0) := v$. Since $v \in D^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$, the definition of $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ implies $\ell_D(w_0) \subseteq \ell(v)$, i.e., φ is a homomorphism from $\mathcal{G}(D)$ into $\mathcal{G}(\mathcal{A})$.

$\mathsf{depth}(D) > 0$, i.e., $D = P_1 \sqcap \ldots \sqcap P_n \sqcap \exists r_1.D_1 \sqcap \ldots \sqcap \exists r_m.D_m$. As for $\mathsf{depth}(D) =)$ we get $\{P_1, \ldots, P_n\} \subseteq \ell(v)$. Now, $v \in D^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ implies that, for each $1 \leq j \leq m$, there exists a node $v_i \in V$ such that $(v, v_i) \in r^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ and $v_i \in D_m^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$. Let w_i denote the r_i-successor of w_0 in $\mathcal{G}(D)$ with $D_i = \mathcal{G}(D)(w_i)$, $1 \leq i \leq m$. By induction, there exist homomorphisms $\varphi_{w_1}, \ldots, \varphi_{w_m}$ from $\mathcal{G}(D)(w_i)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi_{w_i}(w_i) = v_i$ for all $1 \leq i \leq m$. Define φ by

$$\varphi := \{w_0 \mapsto v\} \cup \bigcup_{w_0 r w \in E_D} \varphi_w.$$

Since for each $w_0rw \in E_D$ there exists an $j \in \{1, \dots, m\}$ such that $w = w_j$, φ is well-defined, and by construction, it is a homomorphism from $\mathcal{G}(D)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = v$. □

Whether there exists a homomorphism from a tree into a graph can be decided in polynomial time [8, 12]. Since $\mathcal{G}(C)$ and $\mathcal{G}(\mathcal{A})$ can be computed in polynomial time, we get

Proposition 16 *The instance problem for $\mathcal{EL}$ can be decided in polynomial time.*

3.2 Computing the k-approximations in $\mathcal{EL}$

In this section, we will show that, for an $\mathcal{EL}$ABox $\mathcal{A}$ and an individual $a \in \mathsf{Ind}(abox)$, the k-approximation of a w.r.t. $\mathcal{A}$ always exists and can be effectively computed. The algorithm computing $\mathsf{msc}_{k,\mathcal{A}}(a)$ introduced below works as follows: First, the description graph $\mathcal{G}(\mathcal{A})$ is unraveled into a tree $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ with root a, a finite branching factor, but possibly infinite long paths. Truncating all paths of length $\leq k$ then yields an $\mathcal{EL}$-description tree $\mathcal{T}_k(a, \mathcal{A})$ of depth $\leq k$. Using the characterization of subsumption introduced above, it is easy to show that the $\mathcal{EL}$-concept description $C_{\mathcal{T}_k(\mathcal{G}(\mathcal{A}))}$ is equivalent to $\mathsf{msc}_k(a, \mathcal{A})$. In case that $\mathcal{A}$ is acyclic, $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is an $\mathcal{EL}$-description tree and $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ yields the msc of a w.r.t. $\mathcal{A}$.

For the definition of the trees $\mathcal{T}(a, \mathcal{A})$ and $\mathcal{T}_k(a, \mathcal{A})$, we need the following notions: For an $\mathcal{EL}$-description graph $\mathcal{G} = (V, E, \ell)$, $p = v_0r_1v_1r_2 \dots r_nv_n$ is a *path from* v_0 *to* v_n of *length* $|p| = n$, if $v_{i-1}r_iv_i \in E$ for all $1 \leq i \leq n$. The path p will also be denoted as $r_1 \cdots r_n$*-path from* v_0 *to* v_n, and the node v_n as $r_1 \cdots r_n$*-successor of* v_0, whereby each node is assumed to be an ε-successor of itself. The path p contains a *cycle*, if $v_i = v_j$ for two indices i, j with $0 \leq i < j \leq n$. A node v is *reachable* from v_0, if there exists a path from v_0 to v.

Definition 17 (Tree of a w.r.t. $\mathcal{G}$ (and k)) *Let $\mathcal{G} = (V, E, \ell)$ and $a \in V$. The* tree $\mathcal{T}(a, \mathcal{G})$ of a w.r.t. $\mathcal{A}$ *is defined by* $\mathcal{T}(a, \mathcal{G}) := (V^t, E^t, a, \ell^t)$ *with*

- $V^t := \{ar_1v_1r_2 \dots r_nv_n \mid ar_1v_1r_2 \dots r_nv_n$ *is a path from* a *to* v_n *in* $\mathcal{G}\}$,
- $E^t := \{prq \mid p, q \in V^t$ *and* $p = ar_1v_1r_2 \dots r_nv_n$ *and* $q = ar_1v_1r_2 \dots r_nv_nrw\}$,
- $\ell^t(p) := \ell(v)$ *if* $p = ar_1v_1r_2 \dots r_nv$.

For a nonnegative integer $k \in \mathbb{N}$ the tree $\mathcal{T}_k(a, \mathcal{G})$ of a w.r.t. $\mathcal{G}$ and k *is defined by* $\mathcal{T}_k(a, \mathcal{G}) := (V_k^t, E_k^t, a, \ell_k^t)$ *with*

- $V_k^t := \{p \in V^t \mid |p| \leq k\}$,
- $E_k^t := E^t \cap (V_k^t \times N_R \times V_k^t)$, *and*
- $\ell_k^t(p) := \ell^t(p)$ *if* $p \in V_k^t$.

Since by definition $\mathcal{EL}$-description graphs are finite, $\mathcal{T}(a,\mathcal{G})$ has a finite branching factor, i.e., $\mathcal{T}(a,\mathcal{G})$ is infinite if and only if $\mathcal{T}(a,\mathcal{G})$ contains a path of infinite length. This is the case if and only if there exists a path p in $\mathcal{G}$ such that p contains a cycle. By definition, $\mathcal{T}_k(a,\mathcal{G})$ is an $\mathcal{EL}$-description tree of depth $\leq k$. If $\mathcal{G}$ is acyclic, i.e., $\mathcal{G}$ does not exist a path p in $\mathcal{G}$ containing a cycle, $\mathcal{T}(a,\mathcal{G})$ is finite and thus, an $\mathcal{EL}$-description tree. Now, the following characterization of the (k-approximation of the) MSC is based on these trees, whereby the $\mathcal{EL}$-description graph is obtained from an $\mathcal{EL}$-ABox $\mathcal{A}$, and the root is assumed to be an individual from $\mathcal{A}$. An $\mathcal{EL}$-ABox $\mathcal{A}$ is *acyclic*, of $\mathcal{G}(\mathcal{A})$ is acyclic.

Theorem 18 *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$. Then, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is the k-approximation of a w.r.t. $\mathcal{A}$. If $\mathcal{A}$ is acyclic, the $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ is the Most Specific Concept of a w.r.t. $\mathcal{A}$.*

Proof: Let $\mathcal{G}(\mathcal{A}) = (V_\mathcal{A}, E_\mathcal{A}, \ell_\mathcal{A})$ and $\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})) = (V_k, E_k, a, \ell_k)$.

We first show $a \in_\mathcal{A} C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$. Let φ be the mapping obtained from mapping each path $p \in V_k$ onto the last node v occurring in p. It is easy to see that φ yields a homomorphism from $\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(a) = a$. Since, for $\mathcal{EL}$-description trees $\mathcal{G}$, it is $\mathcal{G} = \mathcal{G}(C_\mathcal{G})$ up to renaming nodes, Theorem 13 yields $a \in_\mathcal{A} C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$.

It remains to show that for all $\mathcal{EL}$-concept descriptions C with $a \in_\mathcal{A} C$ and $\mathsf{depth}(C) \leq k$ also $C_{\mathcal{T}_k(a,\mathcal{A}))} \sqsubseteq C$. Let C be such an $\mathcal{EL}$-concept description. By Theorem 13 we get that there exists a homomorphism φ from $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$. Using φ, we define a homomorphism ψ from $\mathcal{G}(C)$ into $\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))$. Then Theorem 4 implies $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))} \sqsubseteq C$.

For $v \in V_C$, let $v_0 r_1 v_1 r_2 \cdots r_{n-1} v_{n-1} r_n v$ be the unique (!) path from v_0 to v in $\mathcal{G}(C)$. Define $p(v) := \varphi(v_0) r_1 \varphi(v_1) r_2 \cdots r_{n-1} \varphi(v_{n-1}) r_n \varphi(v)$. Since φ is a homomorphism from $\mathcal{G}(C)$ into $G(\mathcal{A})$, $p(v)$ is well-defined and yields a path of length $n \leq k$ from $\varphi(v_0)$ to $\varphi(v)$ in $\mathcal{G}(\mathcal{A})$. It is easy to see that the mapping $\psi : V_C \longrightarrow V_k$ defined by $\psi(v) := p(v)$ yields a homomorphism from $\mathcal{G}(C)$ into $\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))$.

In case that $\mathcal{A}$ is acyclic, $\mathcal{T}(a,\mathcal{A})$ is an $\mathcal{EL}$-description tree. As for $\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))$ it follows that $a \in_\mathcal{A} C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ and $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))} \sqsubseteq C$ for all $\mathcal{EL}$-concept descriptions C with $a \in_\mathcal{A} C$. Thus, $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ is the MSC of a w.r.t. $\mathcal{A}$. □

Since, for an arbitrary $\mathcal{EL}$-ABox $\mathcal{A}$, $\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))$ and, for an acyclic $\mathcal{EL}$-ABox $\mathcal{A}$, $\mathcal{T}(a,\mathcal{G}(\mathcal{A}))$ can be effectively computed, we get the desired result as an easy consequence of Theorem 18.

Corollary 19 *For an $\mathcal{EL}$-ABox $\mathcal{A}$, an individual $a \in \mathsf{Ind}(\mathcal{A})$, and a nonnegative integer $k \in \mathbb{N}$, the k-approximation of a w.r.t. $\mathcal{A}$ always exist and can be effectively computed.*

If $\mathcal{A}$ is acyclic, the MSC of a w.r.t. $\mathcal{A}$ always exists and can be effectively computed.

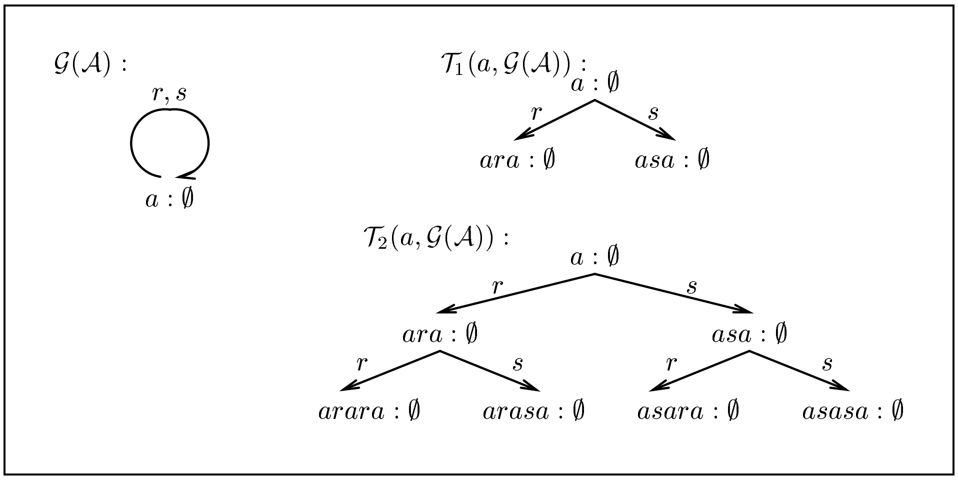

Figure 3: The $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ and the trees of a w.r.t. $\mathcal{A}$ and k, $k = 1, 2$.

3.3 On the complexity of the (k-approximation of the) MSC in $\mathcal{EL}$

Talking about the complexity of the (k-approximation of the) MSC in $\mathcal{EL}$ means talking about the size of $\mathsf{msc}_k(a, \mathcal{A})$ and $\mathsf{msc}(a, \mathcal{A})$ w.r.t. k and the size of $\mathcal{A}$. The *size* of an ABox is defined by

$$|\mathcal{A}| := |\mathsf{Ind}(\mathcal{A})| = |\{(a, b) : r \mid (a, b) : r \in \mathcal{A}\}| + \sum_{a:C \in \mathcal{A}} |C|,$$

where the size $|C|$ of C is defined as the sum of the number of occurences of concept names, role names, and constructors in C.

Example 20 *Let $\mathcal{A} = \{(a, a) : r, (a, a) : s\}$. The $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ as well as the el-description trees $\mathcal{T}_1(a, \mathcal{G}(\mathcal{A}))$ and $\mathcal{T}_2(a, \mathcal{G}(\mathcal{A}))$ are depicted in Figure 3. It is easy to see that $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ yields a full binary tree of depth k where*

- *each node is labeled with the empty set, and*
- *each node except the leaves has exactly one r- and exactly one s-successor.*

By Theorem 18, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is the k-approximation of a w.r.t. $\mathcal{A}$. The size of $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is exponential in the size of $\mathcal{A}$ and k. Analogously to the proofs on the complexity of the LCS of n $\mathcal{EL}$-concept descriptions given in [2] one can show that there does not exist an $\mathcal{EL}$-concept description C which is equivalent to but smaller than $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$.

The above example shows that the k-approximation of a w.r.t. $\mathcal{A}$ can be of exponential size w.r.t. the parameter k, even if $\mathcal{A}$ is of constant size. The following example illustrates that also the MSC of a w.r.t. $\mathcal{A}$ can be of exponential size w.r.t. the size of $\mathcal{A}$.

Example 21 *For $n \geq 1$, define*

$$\mathcal{A}_n := \{(a_i, a_{i+1}) : r, (a_i, a_{i+1}) : s \mid 1 \leq i < n\}.$$

Obviously, $\mathcal{A}_n$ is acyclic, and the size of $\mathcal{A}_n$ is linear in n. By Theorem 18, $C_{\mathcal{T}(a_1,\mathcal{A}_n)}$ is the MSC of a_1 w.r.t. $\mathcal{A}_n$. It is easy to see that, for each n, $\mathcal{T}(a_1, \mathcal{A}_n)$ coincides with the tree $\mathcal{T}_n(a, \mathcal{G}(\mathcal{A}))$ obtained in Example 20. As before we get that

- *$C_{\mathcal{T}(a_1,\mathcal{G}(\mathcal{A}))}$ is of exponential size w.r.t. n, i.e., w.r.t. the size of $\mathcal{A}_n$; and*
- *there does not exist an $\mathcal{EL}$-concept description C equivalent to but smaller then $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$.*

On the one hand, the above examples show that the size of the (approximation of the) MSC can be of exponential size. On the other hand, the size of $\mathsf{msc}_k(a, \mathcal{A})$ and $\mathsf{msc}(a, \mathcal{A})$ can be exponentially bounded: The size of $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ (and hence of $\mathsf{msc}_k(a, \mathcal{A})$) can be exponentially bounded with parameter k: The branching factor of this tree is bounded by the number of role assertions in $\mathcal{A}$ and hence less or equal than $|\mathcal{A}| =: m_0$, where, w.l.o.g., $|\mathcal{A}| > 1$. Since the depth is $\leq k$, the number of nodes is bounded by $\sum_{i=0}^{k} m_0^i = (m_0^{k+1} - 1)/(m_0 - 1)$. Each node label contains at most as many concept names as occurring in $\mathcal{A}$, i.e., each node label has size $\leq |\mathcal{A}|$. Thus, the size of $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ is less or equal than m_0^{k+2}. For an acyclic $\mathcal{EL}$-ABox, also the depth of $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is bounded by the number of role assertions in $\mathcal{A}$. So the size of $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ (and hence of $\mathsf{msc}(a, \mathcal{A})$) is exponentially bounded in the size of $\mathcal{A}$.

Proposition 22 *For an $\mathcal{EL}$-ABox $\mathcal{A}$, $a \in \mathsf{Ind}(\mathcal{A})$, and a nonnegative integer $k \in \mathbb{N}$, the size of*

1. *the k-approximation $\mathsf{msc}_{\mathcal{A},k}(a)$ of a w.r.t. $\mathcal{A}$ is (at most) exponential in k and the size of $\mathcal{A}$; and*
2. *the most specific concept $\mathsf{msc}_{\mathcal{A}}(a)$ of a w.r.t. $\mathcal{A}$ is (at most) exponential in the size of $\mathcal{A}$.*

4 Most Specific Concepts in $\mathcal{EL}_\neg$

The results for $\mathcal{EL}$ obtained in the previous section cannot just be transfered to $\mathcal{EL}_\neg$: First, we have to cope with inconsistent ABoxes as a special case. But even for consistent $\mathcal{EL}_\neg$-ABoxes, the characterization of instance in $\mathcal{EL}$ is not complete, and hence, the algorithm computing the (approximation of the)

MSC in $\mathcal{EL}$ does not yield the (approximation of the) MSC in $\mathcal{EL}_\neg$. In Section 4.1, however, we will fix the problems with the characterization of instance and introduce a sound and complete characterization based on so-called primitive completions. Again using this characterization of instance, we present an algorithm computing the (approximation of the) MSC in Section 4.2.

4.1 The characterization of instance in $\mathcal{EL}_\neg$

First, we show that consistency of $\mathcal{EL}_\neg$-ABoxes can be decided in polynomial time. Therefore, we have to adapt the notions to $\mathcal{EL}_\neg$: *$\mathcal{EL}_\neg$-description trees* and *$\mathcal{EL}_\neg$-description graphs* are defined as for $\mathcal{EL}$ except that nodes are labeled with subsets of the set $N_C \cup \{\neg P \mid P \in N_C\}$. The $\mathcal{EL}_\neg$-description tree $\mathcal{G}(C)$ of an $\mathcal{EL}_\neg$-concept description C is defined as for $\mathcal{EL}$, and also the $\mathcal{EL}_\neg$-description graph $\mathcal{G}(\mathcal{A})$ of an $\mathcal{EL}_\neg$-ABox $\mathcal{A}$. Finally, the notion of the canonical interpretation $\mathcal{I}(\mathcal{G})$ remains unchanged, i.e., negated concept names occurring in node labels do not influence the interpretation of concept names by $P^{\mathcal{I}(\mathcal{G})} := \{v \mid P \in \ell(v)\}$.

Lemma 23 *An $\mathcal{EL}_\neg$-ABox $\mathcal{A}$ is inconsistent iff there exists an individual $a \in \mathsf{Ind}(\mathcal{A})$ with $C_a := \bigsqcap_{a:D\in\mathcal{A}} D \equiv_\perp$.*

Proof: Let $\mathcal{A}$ be an $\mathcal{EL}_\neg$-ABox. If there exists an individual $a \in \mathsf{Ind}(\mathcal{A})$ with $C_a \equiv \perp$, then obviously $\mathcal{A}$ is inconsistent. Conversely, assume $C_a \not\equiv \perp$ for all $a \in \mathsf{Ind}(\mathcal{A})$. We show that the canonical interpretation $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ is a model of $\mathcal{A}$.

It is easy to see that an $\mathcal{EL}_\neg$-concept description C is inconsistent iff there exists a node $v \in V_C$ in the $\mathcal{EL}_\neg$-description tree $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ with $\{P, \neg P\} \subseteq \ell_C(v)$ for some $P \in N_C$. (Otherwise, the canonical interpretation $\mathcal{I}(\mathcal{G}(C))$ yields a model of C with $v_0 \in C^{\mathcal{I}(\mathcal{G}(C))}$.)

Now, the assumption $C_a \not\equiv \perp$ for all $a \in \mathsf{Ind}(\mathcal{A})$ implies that there does not exist a node $v \in V$ in $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$ with $\{P, \neg P\} \subseteq \ell(v)$ for some $P \in N_C$. We have to show that $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ satisfies each assertion in $\mathcal{A}$.

By definition, $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ satisfies each role assertion in $\mathcal{A}$. Let $a : D \in \mathcal{A}$ and $D = Q_1 \sqcap \ldots \sqcap Q_n \sqcap \exists r_1.D_1 \sqcap \ldots \sqcap \exists r_m.D_m$ with $Q_i \in N_C \cup \{\neg P \mid P \in N_C\}$ for all $1 \leq i \leq n$. We show

1. $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in Q_i^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq i \leq n$, and
2. $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in (\exists r_j.D_j)^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq j \leq m$.

Ad (1): Assume $Q_i \in N_C$. By definition, it is $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in Q_i^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$. Assume $Q_i = \neg P$ for some $P \in N_C$. Then $P \notin \ell(a)$, i.e., there does not exist a concept assertion $a : D' \in \mathcal{A}$ such that P occurs in the conjunction on the top-level of D'. By definition, $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \notin P^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$, and hence $a^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \in Q_i^{\mathcal{I}(\mathcal{G}(\mathcal{A}))}$ for all $1 \leq i \leq n$.

Ad (2): Since $C_a \not\equiv \perp$, each D_j is consistent. Then the claim follows as in the proof of Lemma 15. $\square$

As an easy consequence of Lemma 23 and its proof, respectively, we get the following propositions.

Proposition 24 *Consistency of $\mathcal{EL}_\neg$-ABoxes can be decided in polynomial time.*

Lemma 25 *For a consistent $\mathcal{EL}_\neg$-ABox $\mathcal{A}$, the canonical interpretation $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ is a model of $\mathcal{A}$.*

Inconsistent ABoxes can be considered as a special case in the characterization of instance relationships in $\mathcal{EL}_\neg$: If $\mathcal{A}$ is inconsistent, then trivially $a \in_\mathcal{A} C$ for all $\mathcal{EL}_\neg$-concept descriptions C. Restricting the characterization of instance in $\mathcal{EL}$ to consistent $\mathcal{EL}_\neg$-ABoxes, however, does not yield a sound and complete characterization of instance in $\mathcal{EL}_\neg$ (unless P=NP): if $a \in_\mathcal{A} C$ could be decided by testing whether there exists a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$, then by Proposition 16 and Proposition 24 instance in $\mathcal{EL}_\neg$ would be decidable in polynomial time. This would yield a contradiction to the complexity results in [14], though. A thorough investigation of the results in [14] reveals that the instance problem in $\mathcal{EL}_\neg$ is coNP-hard.

The following example is an abstract version of an example given in [14] showing that the naive algorithm described above is not complete for $\mathcal{EL}_\neg$.

Example 26 *Consider the $\mathcal{EL}_\neg$-concept description $C = P \sqcap \exists r.(P \sqcap \exists r.\neg P)$ and the $\mathcal{EL}_\neg$-ABox*

$$\mathcal{A} = \{a : P, b_1 : P, b_3 : \neg P, (a, b_1) : r, (a, b_2) : r, (b_1, b_2) : r, (b_2, b_3) : r\}.$$

$\mathcal{G}(\mathcal{A})$ and $\mathcal{G}(C)$ are depicted in Figure 4. Obviously, there does not exist a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = a$, because neither $\{\neg P\} \subseteq \ell(b_1)$ nor $\{\neg P\} \subseteq \ell(b_2)$. For each model $\mathcal{I}$ of $\mathcal{A}$, however, either $b_2^\mathcal{I} \in P^\mathcal{I}$ or $b_2^\mathcal{I} \in (\neg P)^\mathcal{I}$, and hence, $a^\mathcal{I} \in C^\mathcal{I}$. So a is an instance of C w.r.t. $\mathcal{A}$ though there does not exist a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = a$.

The reason for the problem illustrated in the above example is that, in general, for the individuals in the ABox it is not fixed whether they are instances of a given concept name or not. This problem can be solved by considering the set of so-called *primitive completions* instead of the single $\mathcal{EL}_\neg$-description graph $\mathcal{G}(\mathcal{A})$. Intuitively, a primitive completion of an $\mathcal{EL}_\neg$-description graph $\mathcal{G}(\mathcal{A})$ is an $\mathcal{EL}_\neg$-description graph $\mathcal{G}(\mathcal{A})^*$ with the same set of nodes and edges, but whose node labels are complete in the following sense: for each node v in $\mathcal{G}(\mathcal{A})^*$ and each $P \in N_C$, either P occurs in the label of v, or $\neg P$.

Definition 27 (Primitive completion) *Let $\mathcal{G} = (V, E, \ell)$ be an $\mathcal{EL}_\neg$-description graph and $N_C^* := \{P \in N_C \mid$ exists $v \in V$ with $P \in \ell(v)$ or $\neg P \in \ell(v)\}$. An $\mathcal{EL}_\neg$-description graph $\mathcal{G}^* = (V, E, \ell^*)$ is a* primitive completion *of $\mathcal{G}$ if, for all $v \in V$,*

1. *$\ell(v) \subseteq \ell^*(v)$, and*

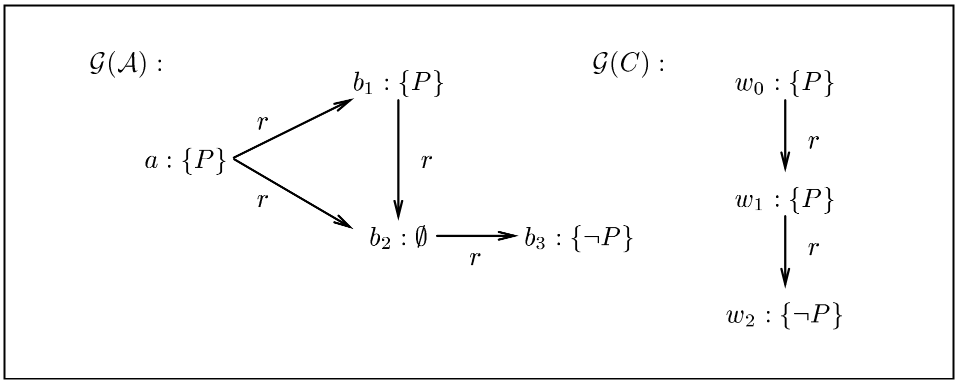

Figure 4: An example for the incompleteness of the characterization of instance in $\mathcal{EL}$ for $\mathcal{EL}_\neg$.

2. *for all concept names $P \in N_C^*$ either $P \in \ell^*(v)$, or $\neg P \in \ell^*(v)$.*

The sound and complete characterization of instance in $\mathcal{EL}_\neg$ is formalized in the following

Theorem 28 *Let $\mathcal{A}$ be an $\mathcal{EL}_\neg$-ABox with $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$ and C an $\mathcal{EL}_\neg$-concept description with $\mathcal{G}(C) = (V_C, E_C, w_0, \ell_C)$. For $a \in \mathsf{Ind}(\mathcal{A})$, it is $a \in_\mathcal{A} C$ iff $\mathcal{A}$ is inconsistent or for each primitive completion $\mathcal{G}(\mathcal{A})^*$ of $\mathcal{G}(\mathcal{A})$, there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$ with $\varphi(w_0) = a$.*

Proof of the if-direction: For the characterization of instance in $\mathcal{EL}$ (see Theorem 13, the proof of the if-direction could be obtained trivially as a special case of the soundness result given for $\mathcal{ALE}$ in Section 5. For $\mathcal{EL}_\neg$, however, things are not that easy: Since for the if-direction we only assume that there exist homomorphisms from $\mathcal{G}(C)$ into *primitive completions* of $\mathcal{G}(\mathcal{A})$, and since a primitive completion of $\mathcal{G}(\mathcal{A})$ in general does not coincide with $\mathcal{G}(\mathcal{A})$, the preconditions of Lemma 36 are not satisfied.

The idea underlying the proof given below is as follows: For a consistent $\mathcal{EL}_\neg$-ABox $\mathcal{A}$ (for inconsistent ABoxes nothing has to be shown) and a model $\mathcal{I} = (\Delta_\mathcal{I}, \cdot^\mathcal{I})$ of $\mathcal{A}$, we first define a mapping $\psi : V \longrightarrow \Delta_\mathcal{I}$ with

1. $\psi(b) = b^\mathcal{I}$ for all $b \in \mathsf{Ind}(\mathcal{A})$,
2. $(\psi(v), \psi(w)) \in r^\mathcal{I}$ for all $vrw \in E$, and
3. $\psi(v) \in Q^\mathcal{I}$ for all $Q \in \ell(v)$ and $v \in V$.

Using this mapping, we then define a primitive completion $\mathcal{G}(\mathcal{A})^*_\psi = (V, E, \ell^*_\psi)$ in such a way that ψ also satisfies condition

4. $\psi(v) \in Q^\mathcal{I}$ for all $Q \in \ell^*_\psi(v)$ and $v \in V$.

By assumption, there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*_\psi$. By induction on the depth of $\mathcal{G}(C)(w)$, we finally show that for all $w \in V_C$

$$\psi(\varphi(w)) \in C_{\mathcal{G}(C)(w)}{}^{\mathcal{I}}. \tag{1}$$

Since $\varphi(w_0) = a$ and $\psi(a) = a^{\mathcal{I}}$ and $C \equiv C_{\mathcal{G}(C)(w_0)}$, this implies $a^{\mathcal{I}} \in C^{\mathcal{I}}$.

The definition of the mapping ψ: Let $a \in \mathsf{Ind}(\mathcal{A})$, $C_a := \bigsqcap_{a:D\in\mathcal{A}} D$, and $\mathcal{G}(C_a) = (V_a, E_a, a, \ell_a)$. For $v \in V_a$, we define $\psi(v)$ by induction on the length λ of the unique (!) path from a to v in $\mathcal{G}(C_a)$ such that in addition $\psi(v) \in C_{\mathcal{G}(C_a)(v)}{}^{\mathcal{I}}$.

$\lambda = 0$: Then $v = a$. Define $\psi(a) := a^{\mathcal{I}}$. Since $\mathcal{I} \models \mathcal{A}$, we get $a^{\mathcal{I}} \in C_a^{\mathcal{I}}$. In particular, $a^{\mathcal{I}} \in Q^{\mathcal{I}}$ for all $Q \in \ell(a)$. Thus, conditions (1) and (3) are satisfied for all $a \in \mathsf{Ind}(\mathcal{A})$.

$\lambda > 0$: Then there exists a unique edge of the form $v'rv \in E_a$. By induction, $\psi(v')$ is already defined, and it is $\psi(v') \in C_{\mathcal{G}(C_a)(v')}{}^{\mathcal{I}}$. Since $v'rv \in E_a$, there exists an existential restriction of the form $\exists r.C_{\mathcal{G}(C_a)(v)}$ on the top-level of $C_{\mathcal{G}(C_a)(v')}$. Now, $\psi(v') \in C_{\mathcal{G}(C_a)(v')}{}^{\mathcal{I}}$ implies that there exists an $\alpha \in \Delta_{\mathcal{I}}$ such that $(\psi(v'), \alpha) \in R^{\mathcal{I}}$ and $\alpha \in C_{\mathcal{G}(C_a)(v)}{}^{\mathcal{I}}$. Define $\psi(v) := \alpha$. Since $\alpha \in C_{\mathcal{G}(C_a)(v)}{}^{\mathcal{I}}$ and $\ell(v) = \ell_a(v)$, we get $\alpha \in Q^{\mathcal{I}}$ for all $Q \in \ell(v)$, i.e., condition (3) is satisfied for v.

By construction, ψ satisfies the conditions (1)–(3).

The definition of the primitive completion $\mathcal{G}(\mathcal{A})^*_\psi$: Let N_C^* be the set of all concept names occurring in $\mathcal{G}(\mathcal{A})$. Define $\mathcal{G}(\ T)^*_\psi := (V, E, \ell^*_\psi)$ by

$$\ell^*_\psi(v) := \{P \in N_C^* \mid \psi(v) \in P^{\mathcal{I}}\} \cup \{\neg P \mid P \in N_C^* \text{ and } \psi(v) \notin P^{\mathcal{I}}\}.$$

By condition (3) we get $\ell(v) \subseteq \ell^*_\psi(v)$ for all $v \in V$. Thus, $\mathcal{G}(\mathcal{A})^*_\psi$ is a primitive completion of $\mathcal{G}(\mathcal{A})$ that, by definition, satisfies condition (4).

Now, the precondition of the if-direction yields a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*_\psi$ with $\varphi(w_0) = a$. We show Property (1) by induction on the depth of $\mathcal{G}(C)(w)$:

$\mathsf{depth}(\mathcal{G}(C)(w)) = 0$: Then $C_{\mathcal{G}(C)(w)} = \bigsqcap_{Q\in\ell_C(w)} Q$.

Since $\ell_C(w) \subseteq \ell^*_\psi(\varphi(w))$, condition (4) implies $\psi(\varphi(w)) \in Q^{\mathcal{I}}$ for all $Q \in \ell_C(w)$. Hence, $\psi(\varphi(w)) \in C_{\mathcal{G}(C)(w)}{}^{\mathcal{I}}$.

$\mathsf{depth}(\mathcal{G}(C)(w)) > 0$: Then $C_{\mathcal{G}(C)(w)} = \bigsqcap_{Q\in\ell_C(w)} Q \sqcap \bigsqcap_{wrw'\in E_C} \exists r.C_{\mathcal{G}(C)(w')}$.

As before we get $\psi(\varphi(w)) \in Q^{\mathcal{I}}$ for all $Q \in \ell_C(w)$. Let $wrw' \in E_C$. By definition of ψ, and since φ is a homomorphism, we get $(\psi(\varphi(w)), \psi(\varphi(w'))) \in r^{\mathcal{I}}$. By induction, $\psi(\varphi(w')) \in C_{\mathcal{G}(C)(w')}{}^{\mathcal{I}}$, and hence $\psi(\varphi(w)) \in (\exists r.C_{\mathcal{G}(C)(w')})^{\mathcal{I}}$.

Summing up, we get $\psi(\varphi(w)) \in C_{\mathcal{G}(C)(w)}{}^{\mathcal{I}}$.

This completes the proof of the if-direction.

Proof of the only-if-direction: Let $\mathcal{A}$ be a consistent $\mathcal{EL}_\neg$-ABox, $N_C^* := \{P \in N_C \mid \exists v \in V : P \in \ell(v) \vee \neg P \in \ell(v)\}$, and $a \in_{\mathcal{A}} C$. Let $\mathcal{G}(C) = (V_C, E_C, w_0, \ell_C)$, $\mathcal{G}(\mathcal{A})^* = (V, E, \ell^*)$ an arbitrary primitive completion of $\mathcal{G}(\mathcal{A})$, and $\mathcal{I}(\mathcal{G}(\mathcal{A})^*)$ the canonical interpretation induced by $\mathcal{G}(\mathcal{A})^*$.

By Lemma 25 $\mathcal{I}(\mathcal{G}(\mathcal{A}))$ is a model of $\mathcal{A}$. Since

- $r^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} = r^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ for all $r \in N_R$,
- $P^{\mathcal{I}(\mathcal{G}(\mathcal{A}))} \subseteq P^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ for all $P \in N_C$, and
- $\neg P \in \ell(v) \implies P \notin \ell^*(v) \implies v \notin P^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$,

$\mathcal{I}(\mathcal{G}(\mathcal{A})^*)$ is also a model of $\mathcal{A}$. We first show that, for all concept names P occurring in C, $P \in N_C^*$. Assume that there exists a concept name P occurring in C with $P \notin N_C^*$. Let $w \in V_C$ be an $r_1 \ldots r_n$-successor of w_0 in $\mathcal{G}(C)$ with $P \in \ell_C(w)$ or $\neg P \in \ell_C(w)$. Assume $P \in \ell_C(w)$. Then, for each model $\mathcal{I}$ of C and each $\alpha \in C^{\mathcal{I}}$, there exists an $(r_1 \ldots r_n)^{\mathcal{I}}$-successor β of α in $\mathcal{I}$ with $\beta \in P^{\mathcal{I}}$. Since $P \notin N_C^*$, however, $P^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)} = \emptyset$, and hence, $C^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)} = \emptyset$ in contradiction to $a \in_{\mathcal{A}} C$. If $\neg P \in \ell_C(w)$, then for each model $\mathcal{I}$ of C and each $\alpha \in C^{\mathcal{I}}$, there exists an $(r_1 \ldots r_n)^{\mathcal{I}}$-successor β of α in $\mathcal{I}$ with $\beta \notin P^{\mathcal{I}}$. Define $\mathcal{J} := (V, \cdot^{\mathcal{J}})$, where $Q^{\mathcal{J}} := Q^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ for all $Q \in N_C^*$ and $P^{\mathcal{J}} := V$. Then $\mathcal{J}$ is a model of $\mathcal{A}$, because the interpretation of concept names and role names occurring in $\mathcal{A}$ remained unchanged w.r.t. $\mathcal{I}(\mathcal{G}(\mathcal{A})^*)$. But obviously, $(\neg P)^{\mathcal{J}} = \emptyset$, and hence $C^{\mathcal{J}} \neq \emptyset$ in contradiction to $a \in_{\mathcal{A}} C$.

Thus, we have shown that, for a consistent $\mathcal{EL}_\neg$-ABox $\mathcal{A}$, $a \in_{\mathcal{A}} C$ implies $P \in N_C^*$ for all concept names P occurring in C. Now, the only-if-direction is an easy consequence of the following

Claim: If $v \in C^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ and if all concept names occurring in C also occur in $\mathcal{A}$, then there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$ with $\varphi(w_0) = v$.

Proof by induction on $\mathsf{depth}(C)$:

$\mathsf{depth}(C) = 0$: Then $C = Q_1 \sqcap \ldots \sqcap Q_n$, where $Q_i \in N_C \cup \{\neg P \mid P \in N_C\}$.

Define φ by $\varphi(w_0) := v$. We know $v \in C^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$. If $Q_i \in N_C$, then by definition of $\mathcal{I}(\mathcal{G}(\mathcal{A})^*)$, it is $Q_i \in \ell(v)$. If $Q_i = \neg P$ for some $P \in N_C$, $v \in Q_i^{\mathcal{I}(\mathcal{G}(\mathcal{A})^*)}$ implies $P \notin \ell(v)$. The precondition on C yields $P \in N_C^*$. Since $\mathcal{G}(\mathcal{A})^*$ is a primitive completion, we get $\neg P \in \ell(v)$. Summing up, it is $\ell_C(w_0) \subseteq \ell^*(v)$, and hence φ is a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$.

$\mathsf{depth}(C) > 0$: The induction step is shown as for $\mathcal{EL}$.

This completes the proof of the only-if-direction and hence of Theorem 28. □

The characterization of instance in $\mathcal{EL}_\neg$ given in Theorem 28 intuitively leads to the following decision algorithm deciding $a \in_{\mathcal{A}} C$:

If $\mathcal{A}$ is inconsistent, return "yes";
else
non-deterministically compute a primitive completion $\mathcal{G}(\mathcal{A})^*$ of $\mathcal{G}(\mathcal{A})$;
if there does not exist a homomorphism φ from $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ into $\mathcal{G}(\mathcal{A})^*$ with $\varphi(v_0) = a$,
then return "no";
else return "yes".

Whether $\mathcal{A}$ is inconsistent can be decided in polynomial time (Proposition 24). $\mathcal{G}(\mathcal{A})^*$) can be computed in non-deterministic polynomial time and the existence of a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$ can be decided in polynomial time. Thus, we have shown that the instance problem for $\mathcal{EL}_\neg$ is in coNP. Together with the hardness result from [14], we get

Proposition 29 *The instance problem for $\mathcal{EL}_\neg$ is coNP-complete.*

4.2 The approximation of the MSC in $\mathcal{EL}_\neg$

Since the characterization of instance for $\mathcal{EL}$ is not complete for $\mathcal{EL}_\neg$, it is not astonishing that the algorithms described in Section 3.2 computing the approximation of the msc and the msc for acyclic $\mathcal{EL}$-ABoxes, respectively, do not yield the desired results for $\mathcal{EL}_\neg$. For instance, consider the $\mathcal{EL}_\neg$-ABox $\mathcal{A}$ from Example 26 again. We get

$$\begin{array}{rcl} C_{\mathcal{T}(a,\mathcal{A})} & = & P \sqcap \\ & & \exists r.\exists r.(\neg P) \sqcap \\ & & \exists r.(P \sqcap \exists r.\exists r.(\neg P)), \end{array}$$

which is not equivalent to the msc of a w.r.t. $\mathcal{A}$. As we will see in the following, the msc of a w.r.t. $\mathcal{A}$ is given by

$$\mathsf{msc}_{\mathcal{A}}(a) = P \sqcap \exists r.(P \sqcap \exists r.\neg P) \sqcap \exists r.(P \sqcap \exists r.\exists r.\neg P),$$

and it is $\mathsf{msc}_{\mathcal{A}}(a) \sqsubset C_{\mathcal{T}(a,\mathcal{A})}$.

In order to compute the msc or the k-approximation of the msc, respectively, all primitive completions must be taken into account. More precisely, we have to extract the commonalities of all primitive completions. In case of a set of concept descriptions the lcs represents all commonalities of these descriptions. In [3] it has been shown that, for a set $\{C_1, \ldots, C_n\}$ of $\mathcal{ALE}$-concept descriptions, the lcs can be obtained from the product of the description trees corresponding to $C_1, \ldots, C_n$. It has turned out that the same idea also works for the msc: the product of the primitive completions of the $\mathcal{EL}_\neg$-description graph $\mathcal{G}(\mathcal{A})$ yields a graph from which the k-approximation/msc can be obtained just as for $\mathcal{EL}$.

Definition 30 *Let $\{\mathcal{G}_i = (V_i, E_i, \ell_i) \mid 1 \leq i \leq n\}$ be a finite set of $\mathcal{EL}_\neg$-description graphs. The* product $\prod_{1 \leq i \leq n} \mathcal{G}_i$ of $\mathcal{G}_1, \ldots, \mathcal{G}_n$ *is defined as the $\mathcal{EL}_\neg$-description graph (V, E, ℓ) where*

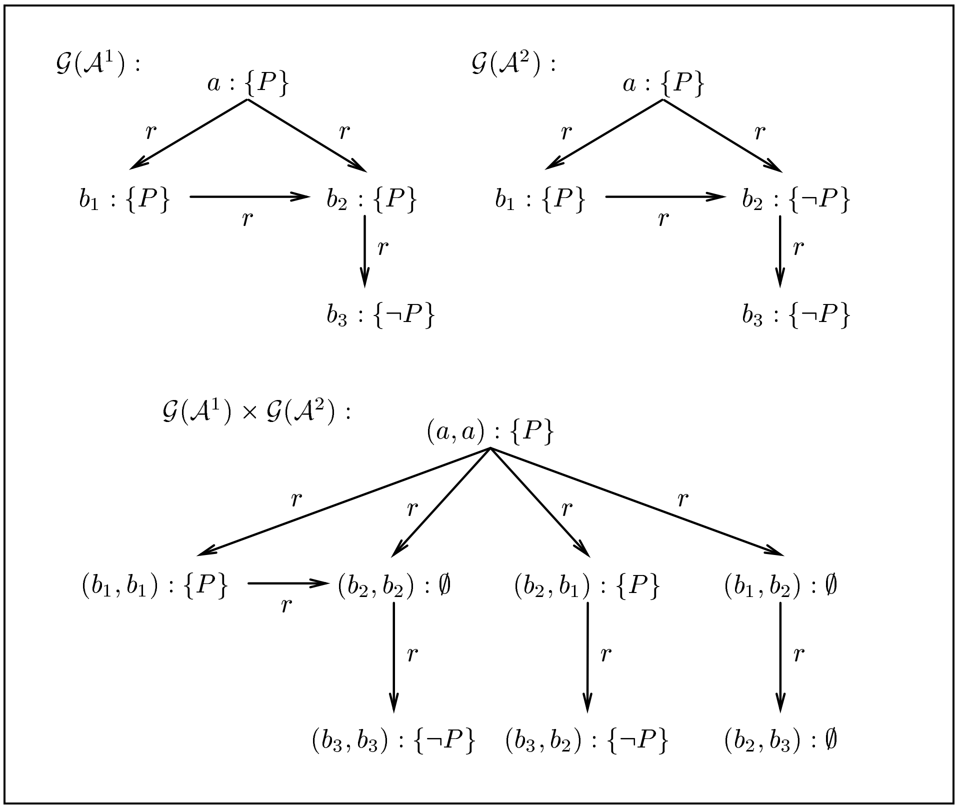

Figure 5: An example for the product of two $\mathcal{EL}_\neg$-description graphs; all nodes that are not reachable from the node (a, a) are neglected.

- $V = V_1 \times \ldots \times V_n$,
- $E = \{(v_1, \ldots, v_n)r(w_1, \ldots, w_n) \mid v_i r w_i \in E_i$ *for all* $1 \leq i \leq n\}$, *and*
- $\ell(v_1, \ldots, v_n) = \bigcap_{1 \leq i \leq n} \ell_i(v_i)$.

The two primitive completions of the $\mathcal{EL}_\neg$-description graph $\mathcal{G}(\mathcal{A})$ from Example 26 are depicted in the upper part of Figure 5; below, the product of $\mathcal{G}(\mathcal{A}^1)$ and $\mathcal{G}(\mathcal{A}^2)$ is depicted, whereby all nodes, which are not reachable from the node (a, a), are neglected. Since we are interested in the tree obtained from unraveling $\mathcal{G}(\mathcal{A}^1) \times \mathcal{G}(\mathcal{A}^2)$ starting with the node (a, a) as its root, this restriction of the product is without loss of generality. In the example, the $\mathcal{EL}_\neg$-description tree $\mathcal{T}((a, a), \mathcal{G}(\mathcal{A}^1) \times \mathcal{G}(\mathcal{A}^2))$ is given by

$$C = P \sqcap \exists r.\exists r.\top \sqcap \exists r.(P \sqcap \exists r.\neg P) \sqcap \exists r.(P \sqcap \exists r.\exists r.\neg P) \sqcap \exists r.\exists r.\neg P.$$

It is easy to see that $C \equiv \mathsf{msc}_{\mathcal{A}}(a)$.

Theorem 31 *Let $\mathcal{A}$ be an $\mathcal{EL}_\neg$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$ a nonnegative integer. If $\mathcal{A}$ is inconsistent, then $\mathsf{msc}_k(a,\mathcal{A}) = \mathsf{msc}(a,\mathcal{A}) = \bot$. Otherwise, let $\{\mathcal{G}(\mathcal{A})^1, \ldots, \mathcal{G}(\mathcal{A})^n\}$ be the set of all primitive completions of $\mathcal{G}(\mathcal{A})$, $\mathcal{G} = (V, E, \ell)$ the product $\prod_{1 \leq i \leq n} \mathcal{G}(\mathcal{A})^i$, and finally, $\boldsymbol{a} \in V$ the n-tuple $(a, \ldots, a)$. Then, $C_{\mathcal{T}_k(\boldsymbol{a},\mathcal{G})}$ is the k-Approximation of a w.r.t. $\mathcal{A}$. If $\mathcal{A}$ is acyclic, then $C_{\mathcal{T}(\boldsymbol{a},\mathcal{G})}$ yields the most specific concept of a w.r.t. $\mathcal{A}$.*

Proof: If $\mathcal{A}$ is inconsistent, then, obviously, $\bot$ is the most specific concept description satisfying the conditions on the k-approximation of a w.r.t. $\mathcal{A}$ and the msc of a w.r.t. $\mathcal{A}$, respectively.

In the sequel, let $\mathcal{A}$ be a consistent $\mathcal{EL}_\neg$-ABox. We show that

1. $a \in_{\mathcal{A}} C_{\mathcal{T}_k(\mathbf{a},\mathcal{G})}$, and
2. for all C with $\mathsf{depth}(C) \leq k$ and $a \in_{\mathcal{A}} C$, it is $C_{\mathcal{T}_k(\mathbf{a},\mathcal{G})} \sqsubseteq C$.

Ad (1): Just as in the proof of Theorem 18, we use the mapping φ that maps each node p in $\mathcal{T}_k(\mathbf{a},\mathcal{G})$ on the last node on the path in $\mathcal{G}$ represented by p. Now, define φ_i as the concatenation $\pi_i \circ \varphi$ of φ and the projection π_i of the i. component of the node $\varphi(p)$. It is easy to see that φ_i yields a homomorphism from $\mathcal{T}_k(\mathbf{a},\mathcal{G})$ into $\mathcal{G}(\mathcal{A})^i$ with $\varphi_i(\mathbf{a}) = a$. Since $\mathcal{T}_k(\mathbf{a},\mathcal{G}))$ coincides with $\mathcal{G}(C_{\mathcal{T}_k(\mathbf{a},\mathcal{G})})$ up to renaming nodes, item (1) is an immediate consequence of Theorem 28.

Ad (2): Let C be an $\mathcal{EL}_\neg$-concept description with $\mathsf{depth}(C) \leq k$ and $a \in_{\mathcal{A}} C$. Let $\mathcal{G}(C) = (V_C, E_C, w_0, \ell_C)$. Since $\mathcal{A}$ is consistent, Theorem 28 implies that there exist homomorphisms φ_i from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^i$ with $\varphi_i(w_0) = a$ for all $1 \leq i \leq n$. Define $\varphi : V_C \longrightarrow V$ by $\varphi(v) := (\varphi_1(v), \ldots, \varphi(v_n))$ for all $v \in V_C$. Since φ_i is a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^i$ for each $1 \leq i \leq n$, φ is well-defined. By induction on $\mathsf{depth}(C)$, one can show that φ is a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}$ with $\varphi(w_0) = \mathbf{a}$. Just as in the proof of Theorem 18, φ can now be used to define a homomorphism ψ from $\mathcal{G}(C)$ into $\mathcal{T}_k(\mathbf{a},\mathcal{G})$. By Theorem 4, this yields $C_{\mathcal{T}_k(\mathcal{G}(\mathcal{A}))} \sqsubseteq C$.

If $\mathcal{A}$ is acyclic, then also the product $\mathcal{G}$ is acyclic. Consequently, $\mathcal{T}(\mathbf{a},\mathcal{G})$ is finite, and $C_{\mathcal{T}(\mathbf{a},\mathcal{G})}$ is an $\mathcal{EL}_\neg$-concept description. Just as for $\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))$ we get

(a) $a \in_{\mathcal{A}} C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ and

(b) $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))} \sqsubseteq C$ for all $\mathcal{EL}_\neg$-concept descriptions C with $a \in_{\mathcal{A}} C$.

So $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ is the msc of a w.r.t. $\mathcal{A}$. □

Corollary 32 *For an $\mathcal{EL}_\neg$-ABox $\mathcal{A}$, an individual $a \in \mathsf{Ind}(\mathcal{A})$, and a nonnegative integer $k \in \mathbb{N}$, the k-approximation $\mathsf{msc}_{\mathcal{A},k}(a)$ of a w.r.t. $\mathcal{A}$ always exists and can effectively be computed.*

If $\mathcal{A}$ is acyclic, the most specific concept $\mathsf{msc}_{\mathcal{A}}(a)$ of a w.r.t. $\mathcal{A}$ always exists and can effectively be computed.

Note that for the effective computation of the k-approximation and the msc, respectively, the restriction to the finite set N_C^* of concept names actually occurring in $\mathcal{A}$ in the definition of primitive completions is crucial. Otherwise, one would have to restrict the result to finite signatures, more precisely: to a finite set of concept names, in order to effectively compute the set of primitive completions and hence, the k-approximation and msc, respectively.

4.3 The complexity of the (k-approximation of the) msc in $\mathcal{EL}_\neg$

The examples given in Section 3.3 are based on ABoxes that do not contain any concept names. Thus, if we consider the ABox $\mathcal{A}$ from Example 20 and the ABoxes $\mathcal{A}_i$, $i \geq 1$, from Example 21 as $\mathcal{EL}_\neg$-ABoxes over $N_R = \{r, s\}$ and $N_C = \emptyset$, then $\mathcal{G}(\mathcal{A})$ and $\mathcal{G}(\mathcal{A}_i)$, respectively, yield the unique primitive of completion of themselves. Thus, the concept description $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ from Example 20 yields the k-approximation of a w.r.t. $\mathcal{A}$ and the concept descriptions $C_{\mathcal{T}(a_1,\mathcal{G}(\mathcal{A}_i))}$ from Example 21 yield the msc of a_1 w.r.t. $\mathcal{A}_i$, $i \geq 1$. Again, there do not exist equivalent but smaller $\mathcal{EL}_\neg$-concept descriptions. Thus, we get

Proposition 33 *For an $\mathcal{EL}_\neg$-ABox $\mathcal{A}$, an individual $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$*

1. *the k-approximation $\mathsf{msc}_{\mathcal{A},k}(a)$ of a w.r.t. $\mathcal{A}$ can be of size exponential in the size of k and the size of $\mathcal{A}$; and*
2. *the most specific concept $\mathsf{msc}_{\mathcal{A}}(a)$ of a w.r.t. $\mathcal{A}$ can be of size exponential in the size of $\mathcal{A}$.*

Searching an upper bound for the size of the k-approximation/msc as for $\mathcal{EL}$ only leads to a double exponential bound: Each primitive completion has size polynomial in the size of $\mathcal{A}$. However, the number of primitive completions can be exponential in the size of $\mathcal{A}$. The size of the trees with root a one obtains from each primitive completion can again be exponential in k and the size of $\mathcal{A}$. Hence, we get a double exponential bound for the size of the product, i.e., the k-approximation/msc. But since the k-approximation/msc has to be embedded into each primitive completion, and since each primitive completion is of polynomial size, we strongly conjecture that the size of the k-approximation/msc can be exponentially bounded. Proving this conjecture remains an open problem in this work, though.

5 Most Specific Concepts in $\mathcal{ALE}$

As already mentioned in Section 1, the characterization of instance in $\mathcal{EL}_\neg$ could not be extended to $\mathcal{ALE}$. Since such a characterization was crucial for the algorithms computing the k-approximation/msc in el and $\mathcal{EL}_\neg$, respectively, no similar algorithms for $\mathcal{ALE}$ can be presented here. However, we will

1. show that, given that N_C and N_R are finite sets, the k-approximation of a w.r.t. an $\mathcal{ALE}$-ABox $\mathcal{A}$ always exists and can be computed;
2. show that the characterization of instance for $\mathcal{EL}$ is still sound for $\mathcal{ALE}$ (cf Lemma 36); and
3. illustrate the main problem caused by value restrictions in the structural characterization of instance in $\mathcal{ALE}$ (cf Example 37).

The soundness result formalized in Lemma 36 yields the basis for an "'approximation of the k-approximation"' (cf Lemma 38), which might lead to already interesting results in some applications. Furthermore, we hope that a thorough investigation of the examples illustrating incompleteness for $\mathcal{ALE}$ of the naive characterization of instance obtained from $\mathcal{EL}$ due to primitive negation and value restrictions, respectively, may give some hints towards a sound and complete characterization of instance for $\mathcal{ALE}$ that again allows for better algorithms computing the k-approximation/msc in $\mathcal{ALE}$.

5.1 Existence and computability of the k-approximation in $\mathcal{ALE}$

Theorem 34 *Let N_C be a finite set of concept names and N_R a finite set of role names. Let $\mathcal{A}$ be an $\mathcal{ALE}$-ABox such that $\mathcal{A}$ contains only names from $N_C \cup N_R$. Then, for a nonnegative integer $k \in \mathbb{N}$ and an individual $a \in \mathsf{Ind}(\mathcal{A})$, the k-approximation of a w.r.t. $\mathcal{A}$ always exists and can effectively be computed.*

Proof: By induction on k it is easy to see that

- for finite sets N_C and N_R, there exist only finitely many different (modulo equivalence) $\mathcal{ALE}$-concept descriptions with depth $\leq k$, i.e., the set $\mathcal{C}_k := \{[C]_{\equiv,k} \mid \mathsf{depth}(C) \leq k\}$ is finite, and
- can be effectively computed,

where $[C]_{\equiv,k} := \{D \mid \mathsf{depth}(D) \leq k \text{ and } C \equiv D\}$.

If $\mathcal{A}$ is inconsistent, it is $\mathsf{msc}_{\mathcal{A}}(a) = \bot$ and nothing has to be shown. Let $\mathcal{A}$ be a consistent $\mathcal{ALE}$-ABox. Define $\mathcal{C}_k(a, \mathcal{A}) := \{[C]_{\equiv,k} \in \mathcal{C}_k \mid a \in_{\mathcal{A}} C\}$. Since $\mathcal{C}_k$ is finite and can be effectively computed, and since the instance problem for $\mathcal{ALE}$ is decidable [14], the set $\mathcal{C}_k(a, \mathcal{A})$ is finite and can be effectively computed. We get

$$\mathsf{msc}_{\mathcal{A},k}(a) \equiv \bigsqcap_{[C]_{\equiv,k} \in \mathcal{C}_k(a,\mathcal{A})} C,$$

where for each class from $\mathcal{C}_k(a, \mathcal{A})$ exactly one representative is considered. Thus, the conjunction is finite and hence, the k-approximation of a w.r.t. $\mathcal{A}$ can effectively be computed. □

For $\mathcal{EL}$ and $\mathcal{EL}_{\neg}$, the proof of existence of the k-approximation could be easily modified to proof existence of the msc of acyclic ABoxes. Unfortunately, we do not get a similar result for $\mathcal{ALE}$. Since it is not clear whether there exists a finite bound for the size of the msc of a w.r.t. an acyclic $\mathcal{ALE}$-ABox, we cannot re-use the idea underlying the proof of Theorem 34 in order to prove that, for acyclic $\mathcal{ALE}$-ABoxes, the msc always exists and can effectively be computed. In fact, existence of the msc for acyclic $\mathcal{ALE}$-ABoxes remains an open problem.

Since in practical applications the underlying sets of concept names and role names are mostly finite, this condition in Theorem 34 should yield no severe restriction from a practical point of view. However, the algorithm one would obtain from the proof of Theorem 34 is obviously very inefficient. Thus the following section is concerned with a more efficient algorithm computing an approximation of the k-approximation that might already be sufficient for some applications.

5.2 An approximation of the k-approximation in $\mathcal{ALE}$

In order to show that the naive adaption of the characterization of instance from $\mathcal{EL}$ to $\mathcal{ALE}$ at least yields a sound characterization of instance for $\mathcal{ALE}$, we need some notations:

An *$\mathcal{ALE}$-description graph* $\mathcal{G} = (V, E, \ell)$ is a labeled graph with

- $E \subseteq V \times N_R \times V \cup V \times \{\forall r \mid r \in N_R\} \times V$ and
- $\ell(v) \subseteq N_C \cup \{\neg P \mid P \in N_C\} \cup \{\bot\}$ for all $v \in V$.

Definition 35 ($\mathcal{ALE}$-description graph of an $\mathcal{ALE}$-ABox) *Let $\mathcal{A}$ be an $\mathcal{ALE}$-ABox. For all $a \in \mathsf{Ind}(\mathcal{A})$) let $C_a := \bigsqcap_{a:D\in\mathcal{A}} D$, if there exists a concept assertion of the form $a : D$ in $\mathcal{A}$; otherwise, let $C_a := \top$. Let $\mathcal{G}(C_a) = (V_a, E_a, a, \ell_a)$ denote the $\mathcal{ALE}$-description trees of C_a, $a \in \mathsf{Ind}(\mathcal{A})$, where w.l.o.g. the sets V_a are assumed to be disjoint.*

The $\mathcal{ALE}$-description graph $\mathcal{G}(\mathcal{A})$ of $\mathcal{A}$ *is defined by $\mathcal{G}(\mathcal{A}) := (V, E, \ell)$ with*

- $V := \bigcup_{a\in\mathsf{Ind}(\mathcal{A})} V_a$,
- $E := \{arb \mid (a, b) : r \in \mathcal{A}\} \cup \bigcup_{a\in\mathsf{Ind}(\mathcal{A})} E_a$, *and*
- $\ell(v) := \ell_a(v)$ *for all* $v \in V_a$.

A homomorphism φ from an $\mathcal{ALE}$-description tree $\mathcal{H} = (V_H, E_H, v_0, \ell_H)$ into an $\mathcal{ALE}$-description graph $\mathcal{G} = (V, E, \ell)$ is a mapping $\varphi : V_H \longrightarrow V$ satisfying the conditions (2) and (3) from Definition 3.

Lemma 36 *Let $\mathcal{A}$ be an $\mathcal{ALE}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$ an individual in $\mathcal{A}$, and C an $\mathcal{ALE}$-concept description. Further, let $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$ denote the $\mathcal{ALE}$-description graph of $\mathcal{A}$ and $\mathcal{G}(C) = (V_C, E_C, v_0, \ell_C)$ the $\mathcal{ALE}$-description tree of C. Then, $a \in_{\mathcal{A}} C$ if there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$.*

Proof: If $\mathcal{A}$ is inconsistent, nothing has to be shown. Let $\mathcal{A}$ be a consistent $\mathcal{ALE}$-ABox and $\mathcal{I}$ a model of $\mathcal{A}$. Let $C_a = \bigsqcap_{a:D\in\mathcal{A}} D$ and $\mathcal{G}(C_a) = (V_a, E_a, a, \ell_a)$. Now, $\mathcal{I} \models \mathcal{A}$ implies $a^{\mathcal{I}} \in C_a^{\mathcal{I}}$. We show $a^{\mathcal{I}} \in C^{\mathcal{I}}$ by induction on $\mathsf{depth}(C)$:

$\mathsf{depth}(C) = 0$: Then $C = Q_1 \sqcap \ldots \sqcap Q_n$ with $Q_i \in N_C \cup \{\neg P \mid P \in N_C\} \cup \{\top, \bot\}$. We show $a^{\mathcal{I}} \in Q_i^{\mathcal{I}}$ for all $1 \leq i \leq n$.

For $Q_i = \top$ nothing has to be shown.

Assume $Q_i = \bot$. This would imply $C_a \equiv \bot$ in contradiction to $\mathcal{I} \models \mathcal{A}$.

Assume $Q_i \in N_C$ or $Q_i = \neg P$ for some $P \in N_C$. Then, $P \in \ell(a)$ or $\neg P \in \ell(a)$. By definition of $\mathcal{G}(C_a)$ and C_a, we get $C_a \sqsubseteq P$ or $C_a \sqsubseteq \neg P$, respectively, and hence $a^{\mathcal{I}} \in P^{\mathcal{I}}$ or $a^{\mathcal{I}} \in (\neg P)^{\mathcal{I}}$.

$\mathsf{depth}(C) > 0$: Then $C = Q_1 \sqcap \ldots \sqcap Q_n \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_m.C_m \sqcap \forall s_1.D_1 \sqcap \ldots \sqcap \forall s_k.D_k$ with $Q_i \in N_C \cup \{\neg P \mid P \in N_C\} \cup \{\top \bot\}$. We show $a^{\mathcal{I}} \in C'^{\mathcal{I}}$ for all conjuncts C' on the top-level of C.

For $C' = Q_i$, the claim follows as for $\mathsf{depth}(C) = 0$.

Let $C' = \forall s_j.D_j$ and $w \in V_C$ the $\forall s_j$-successor of w_0 with $\mathcal{G}(C)(w) = \mathcal{G}(D_j)$. Since φ is a homomorphism, there exists a $\forall s_j$-successor v of a in $\mathcal{G}(\mathcal{A})$ with $\varphi(w) = v$. By definition of $\mathcal{G}(\mathcal{A})$, this node v is the root of a subtree of $\mathcal{G}(C_a)$. In particular, $C_a \sqsubseteq \forall s_j.C_{\mathcal{G}(C_a)(v)}$. Thus, $a^{\mathcal{I}} \in C_a^{\mathcal{I}}$ implies $a^{\mathcal{I}} \in (\forall s_j.C_{\mathcal{G}(C_a)(v)})^{\mathcal{I}}$. Obviously, restricting φ to the nodes in $\mathcal{G}(C)(w)$ yields a homomorphism from $\mathcal{G}(C)(w)$ into $\mathcal{G}(C_a)(v)$. By Theorem 4, it follows $C_{\mathcal{G}(C_a)(v)} \sqsubseteq C_{\mathcal{G}(C)(w)}$, and hence $a^{\mathcal{I}} \in (\forall s_j.D_j)^{\mathcal{I}}$.

Let $C' = \exists r_j.C_j$ and $w \in V_C$ the r_j-successor of w_0 with $\mathcal{G}(C)(w) = \mathcal{G}(C_j)$. Since φ is a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$, there exists an r_j-successor v of a in $\mathcal{G}(\mathcal{A})$ with $\varphi(w) = v$. If $v \notin \mathsf{Ind}(\mathcal{A})$, then y definition of $\mathcal{G}(\mathcal{A})$, this node v is the root of a subtree of $\mathcal{G}(C_a)$. As in the previous case, we get $a^{\mathcal{I}} \in (\exists r_j.C_j)^{\mathcal{I}}$. If $v \in \mathsf{Ind}(\mathcal{A})$, then φ restricted to the nodes in $\mathcal{G}(C)(w)$ yields a homomorphism ψ from $\mathcal{G}(C)(w)$ into $\mathcal{G}(\mathcal{A})$ with $\psi(w) = v$. By induction, we get $v \in_{\mathcal{A}} C_{\mathcal{G}(C)(w)}$. Since $\mathcal{I} \models \mathcal{A}$, it follows $v^{\mathcal{I}} \in C^{\mathcal{I}}_{\mathcal{G}(C)(w)}$, and since $C_j \equiv C_{\mathcal{G}(C)(w)}$, this yields $a^{\mathcal{I}} \in (\exists r_j.C_j)^{\mathcal{I}}$. $\square$

The following example shows that the characterization of instance is incomplete for $\mathcal{ALE}$, even if we take into account all primitive completions.

Example 37 *Consider the $\mathcal{ALE}$-ABox*

$$\begin{aligned} \mathcal{A} \;:=\; & \{a : P, b_1 : P \sqcap \forall s.P \sqcap \exists r.P, b_2 : P \sqcap \exists r.(P \sqcap \exists s : P), \\ & (a, b_1) : r, (a, b_2) : r, (b_1, b_2) : r\}. \end{aligned}$$

The corresponding $\mathcal{ALE}$-description graph is depicted on the left-hand side of Figure 6, and on the right-hand side the $\mathcal{ALE}$-description tree of

$$C = \exists r.(\forall s.P \sqcap \exists r.\exists s.\top).$$

Please note that $\mathcal{G}(\mathcal{A})$ is the unique primitive completion of itself (w.r.t. $N_C = \{P\}$), C is in $\top$-normal form, and all concept descriptions occurring in concept assertions of $\mathcal{A}$ are in $\mathcal{ALE}$-normal form.

However, there does not exist a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = a$: if one maps w_1 onto b_1, then there is no rs-successor of b_1 the

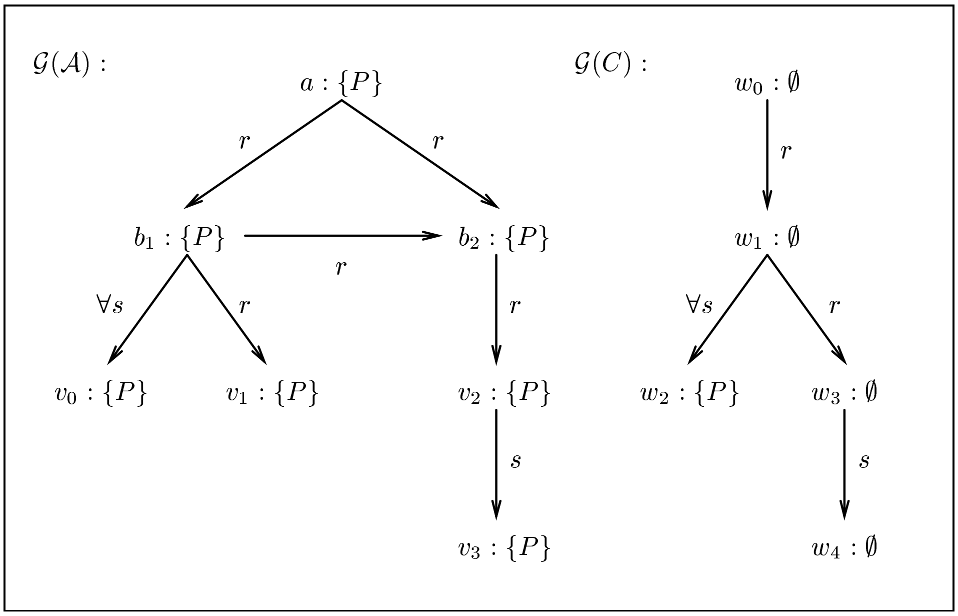

Figure 6: An example illustrating incompleteness of the characterization of instance for $\mathcal{ALE}$.

node w_4 can be mapped onto; otherwise, if one maps w_1 onto b_2, there is no $\forall s$-successor of b_2 the node w_3 can be mapped onto.

Nevertheless $a \in_{\mathcal{A}} C$: For each model $\mathcal{I}$ of $\mathcal{A}$, $b_2^{\mathcal{I}}$ does not have an s-successor or at least one s-successor. In the first case, $b_2^{\mathcal{I}} \in \forall s.P$, and hence $b_2^{\mathcal{I}}$ yields the desired r-successor of $a^{\mathcal{I}}$ in $(\forall s.P \sqcap \exists r.\exists s.\top)^{\mathcal{I}}$. In the second case, it is $b_2^{\mathcal{I}} \in (\exists s.\top)^{\mathcal{I}}$, and hence $b_1^{\mathcal{I}}$ yields the desired r-successor of $a^{\mathcal{I}}$. So for each model $\mathcal{I}$ of $\mathcal{A}$, it is $a^{\mathcal{I}} \in C^{\mathcal{I}}$.

Intuitively, the above example suggests that the notion 'primitive completion' must be extended from (negated) concept names to sets of more complex concept descriptions. So far it is not clear, however, for which set of concept descriptions one has to decide whether an individual is an instance of them or not. Furthermore, it should be noted that the above example does not include primitive negation, i.e., the problem arises just from value restrictions that are not explicitly represented in the ABox. As mentioned before, an appropriate structural characterization of instance for $\mathcal{ALE}$—"appropriate" in the sense that it intuitively leads to a more efficient algorithm for the k-approximation/msc in $\mathcal{ALE}$—remains an open problem within this work.

At the end of this section we show that for an $\mathcal{ALE}$-ABox $\mathcal{A}$ and $a \in \mathsf{Ind}(\mathcal{A})$, $a \in_{\mathcal{A}} C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ for all $k \geq 0$. The trees $\mathcal{T}(a,\mathcal{G}(\mathcal{A}))$ and $\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))$ are defined just as for $\mathcal{EL}$.

Lemma 38 *Let $\mathcal{A}$ be an $\mathcal{ALE}$-ABox. For $a \in \mathsf{Ind}(\mathcal{A})$ and $k \geq 0$, it is $a \in C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$.*

Proof: Just as in the proof of Theorem 18, one can define a mapping φ from $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ into $\mathcal{G}(\mathcal{A})$ such that φ is a homomorphism with $\varphi(a) = a$. Then Lemma 36 implies $a \in C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$. □

In general, the $\mathcal{ALE}$-concept description $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is not the k-approximation of a w.r.t. $\mathcal{A}$. For example, from the ABox $\mathcal{A}$ in Example 37 we get

$$\begin{array}{rcl} C_{\mathcal{T}_4(a,\mathcal{A})} & = & P \sqcap \\ & & \exists r.(P \sqcap \forall s.P \sqcap \exists r.P \sqcap \exists r.(P \sqcap \exists r.(P \sqcap \exists s.P))) \sqcap \\ & & \exists r.(P \sqcap \exists r.(P \sqcap \exists s.P)). \end{array}$$

For the concept description C from Example 37, we have already shown $a \in_{\mathcal{A}} C$. Thus, $a \in_{\mathcal{A}} C \sqcap C_{\mathcal{T}_4(a,\mathcal{A})}$. Since $C_{\mathcal{T}_4(a,\mathcal{A})} \not\sqsubseteq C$, we get $C_{\mathcal{T}_4(a,\mathcal{A})} \sqcap C \sqsubset C_{\mathcal{T}_4(a,\mathcal{A})}$, and hence, $C_{\mathcal{T}_4(a,\mathcal{A})}$ is not equivalent to the 4-approximation of a w.r.t. $\mathcal{A}$.

In order to obtain a rather specific approximation of the k-approximation, as many implicit assertions as possible should be made explicit in $\mathcal{A}$. Formally, this means that the propagation rule

$\mathcal{A} \longrightarrow \mathcal{A} \cup \{b : D\}$, if
  $\{(a, b) : r, a : C\} \subseteq \mathcal{A}$ and
  $\forall r.D$ occurs on the top-level of C and
  there does not exist a concept assertion $b : C'$ in $\mathcal{A}$ such that $C' \sqsubseteq D$

should be applied exhaustively to $\mathcal{A}$ before computing $\mathcal{T}_k(a, \mathcal{A})$. Let $\mathcal{A}^*$ denote the $\mathcal{ALE}$-ABox obtained this way. Then obviously, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}^*))} \sqsubseteq C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$.

6 Conclusion

This work was concerned with the problem of computing most specific concepts, a non-standard inference service needed in different applications, e.g. the bottom-up construction of knowledge bases [3, 4, 13] or learning concept descriptions [7]. Motivated by our application in chemical process engineering [3, 4, 13], we concentrated on DLs with existential restrictions. It has been shown that for $\mathcal{ALE}$ and even the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$ the msc need not exist. As indicated by the results for $\mathcal{ALN}$ [1], an exact characterization of the msc would most probably require cyclic $\mathcal{ALE}$-concept descriptions interpreted by the greatest fixed-point semantics. Since cyclic $\mathcal{ALE}$-concept descriptions and their semantics are not yet well-investigated, we restricted our attention to the approximation of the msc.

Starting with the formal definition of the k-approximation, which has been first introduced in [7], we showed that, for $\mathcal{ALE}$ and a finite signature (N_C, N_R), the k-approximation always exists and can effectively be computed. Unfortunately, we could not give a characterization of instance in $\mathcal{ALE}$ leading to a workable algorithm. Such an algorithm has only been introduced for the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$, where as a by-product, we could show that, for acyclic $\mathcal{EL}$-/$\mathcal{EL}_\neg$-ABoxes, the msc always exists and can effectively be computed. Though we did not get that far for $\mathcal{ALE}$, we have pointed out the main problems encountered when searching for appropriate characterizations of instance, the msc, or its k-approximation. So this work should be a good starting point for further investigations on computing msc in DLs with existential restrictions.

References

[1] F. Baader and R. Küsters. Computing the least common subsumer and the most specific concept in the presence of cyclic $\mathcal{ALN}$-concept descriptions. In O. Herzog and A. Günter, editors, *Proceedings of the 22nd Annual German Conference on Artificial Intelligence (KI'98)*, volume 1504 of *Lecture Notes in Computer Science*, pages 129–140. Springer-Verlag, 1998.

[2] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. LTCS-Report 98-09, LuFG Theoretical Computer Science, RWTH Aachen, Germany, 1998. See http://www-lti.informatik.rwth-aachen.de/Forschung/Reports.html.

[3] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In T. Dean, editor, *Proceedings of the 16th International Joint Conference on Artificial Intelligence 1999 (IJCAI'99)*, pages 96–101. Morgan Kaufmann, 1999.

[4] F. Baader and U. Sattler. Knowledge representation in process engineering. In L. Padgham, E. Franconi, M. Gehrke, D. McGuinness, and P. Patel-Schneider, editors, *Proceedings of the 1996 International Workshop on Description Logic (DL'96)*, pages 74–78. AAAI Press, 1996.

[5] R.J. Brachman, D. McGuinness, P. Patel-Schneider, L. Resnick, and A. Borgida. Living with CLASSIC: When and how to use a KL-ONE-like language. In John F. Sowa, editor, *Principles of Semantic Networks*. Morgan Kaufmann, 1991.

[6] W.W. Cohen, A. Borgida, and H. Hirsh. Computing least common subsumers in description logics. In W. Swartout, editor, *Proceedings of the 10th National Conference on Artificial Intelligence*, pages 754–760. MIT Press, 1992.

[7] W.W. Cohen and H. Hirsh. Learning the CLASSIC description logic: Theoretical and experimental results. In J. Doyle, E. Sandewall, and P. Torasso, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the 4th International Conference (KR'94)*, pages 121–132. Morgan Kaufmann, 1994.

[8] M.R. Garey and D.S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness.* Freeman, 1979.

[9] R. Küsters. Characterizing the semantics of terminological cycles in $\mathcal{ALN}$ using finite automata. In A.G. Cohn, L. Schubert, and S.C. Shapiro, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of 6th International Conference (KR'98)*, pages 499–510. Morgan Kaufmann, 1998.

[10] R. Küsters. *Non-Standard Inference Services in Description Logics.* PhD thesis, RWTH Aachen, 2000.

[11] R. Küsters and A. Borgida. What's in an attribute? Consequences for the least common subsumer. Technical Report DCS-TR-404, Rutgers University, USA, 1999.

[12] S.W. Reyner. An analysis of a good algorithm for the subtree problem. *SIAM Journal of Computing*, 6(4):730–732, December 1977.

[13] U. Sattler. *Terminological knowledge representation systems in a process engineering application.* PhD thesis, RWTH Aachen, 1998. Siehe http://www-lti.informatik.rwth-aachen.de/Forschung/Papers.html.

[14] A. Schaerf. On the complexity of the instance checking problem in concept languages with existential quantification. *Journal of Intelligent Information Systems*, 2:265–278, 1993.