Approximating Most Specific Concepts in Description Logics with Existential Restrictions

Ralf Küsters

Institut für Informatik und
Praktische Mathematik
Christian-Albrechts-Universität zu
Kiel, Germany
kuesters@ti.informatik.uni-kiel.de

Ralf Molitor*

IT Research & Development Group

Swiss Life
Zürich, Switzerland
ralf.molitor@swisslife.ch

Abstract. Computing the most specific concept (msc) is an inference task that allows to abstract from individuals defined in description logic (DL) knowledge bases. For DLs that allow for existential restrictions or number restrictions, however, the msc need not exist unless one allows for cyclic concepts interpreted with the greatest fixed-point semantics. Since such concepts cannot be handled by current DL-systems, we propose to approximate the msc. We show that for the DL $\mathcal{ALE}$, which has concept conjunction, a restricted form of negation, existential restrictions, and value restrictions as constructors, approximations of the msc always exist and can effectively be computed.

1 Introduction

The most specific concept (msc) of an individual b is a concept description that has b as instance and is the least concept description (w.r.t. subsumption) with this property. Roughly speaking, the msc is the concept description that, among all concept descriptions of a given DL, represents b best. Closely related to the msc is the least common subsumer (lcs), which, given concept descriptions $C_1, \ldots, C_n$, is the least concept description (w.r.t. subsumption) subsuming $C_1, \ldots, C_n$. Thus, where the msc generalizes an individual, the lcs generalizes a set of concept descriptions.

In [2–4], the msc (first introduced in [15]) and the lcs (first introduced in [5]) have been proposed to support the bottom-up construction of a knowledge base. The motivation comes from an application in chemical process engineering [17], where the process engineers construct the knowledge base (which consists of descriptions of standard building blocks of process models) as follows: First, they introduce several "typical" examples of a 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

* This work was carried out while the author was still at the LuFG Theoretische Informatik, RWTH Aachen, Germany.

specific description satisfying property (i). The task of computing a concept description satisfying (i) and (ii) can be split into two subtasks: computing the msc of a single individual, and computing the lcs of a given finite number of concepts. The resulting concept description is then presented to the knowledge engineer, who can trim the description according to his needs before adding it to the knowledge base.

The lcs has been thoroughly investigated for (sublanguages of) CLASSIC [5, 2, 12, 11], for DLs allowing for existential restrictions like $\mathcal{ALE}$ [3], and most recently, for $\mathcal{ALEN}$, a DL allowing for both existential and number restrictions [13]. For all these DLs, except for CLASSIC in case attributes are interpreted as total functions [12], it has turned out that the lcs always exists and that it can effectively be computed. Prototypical implementations show that the lcs algorithms behave quite well in practice [7, 4].

For the msc, the situation is not that rosy. For DLs allowing for number restrictions or existential restrictions, the msc does not exist in general. Hence, the first step in the bottom-up construction, namely computing the msc, cannot be performed. In [2], it has been shown that for $\mathcal{ALN}$, a sublanguage of CLASSIC, the existence of the msc can be guaranteed if one allows for cyclic concept descriptions, i.e., concepts with cyclic definitions, interpreted by the greatest fixed-point semantics. Most likely, such concept descriptions would also guarantee the existence of the msc in DLs with existential restrictions. However, current DL-systems, like FaCT [10] and RACE [9], do not support this kind of cyclic concept descriptions: although they allow for cyclic definitions of concepts, these systems do not employ the greatest fixed-point semantics, but descriptive semantics. Consequently, cyclic concept descriptions returned by algorithms computing the msc cannot be processed by these systems.

In this paper, we therefore propose to approximate the msc. Roughly speaking, for some given non-negative integer k, the *k-approximation* of the msc of an individual b is the least concept description (w.r.t. subsumption) among all concept descriptions with b as instance and role depth at most k. That is, the set of potential most specific concepts is restricted to the set of concept descriptions with role depth bounded by k. For (sublanguages of) $\mathcal{ALE}$ we show that k-approximations always exist and that they can effectively be computed. Thus, when replacing "msc" by "k-approximation", the first step of the bottom-up construction can always be carried out. Although the original outcome of this step is only approximated, this might in fact suffice as a first suggestion to the knowledge engineer.

While for full $\mathcal{ALE}$ our k-approximation algorithm is of questionable practical use (since it employs a simple enumeration argument), we propose improved algorithms for the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$ of $\mathcal{ALE}$. ($\mathcal{EL}$ allows for conjunction and existential restrictions, and $\mathcal{EL}_\neg$ additionally allows for a restricted form of negation.) Our approach for computing k-approximations in these sublanguages is based on representing concept descriptions by certain trees and ABoxes by certain (systems of) graphs, and then characterizing instance relationships by homomorphisms from trees into graphs. The k-approximation operation then consists in unraveling the graphs into trees and translating them back into con-

Table 1. Syntax and semantics of concept descriptions.

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

cept descriptions. In case the unraveling yields finite trees, the corresponding concept descriptions are "exact" most specific concepts, showing that in this case the msc exists. Otherwise, pruning the infinite trees on level k yields k-approximations of the most specific concepts.

The outline of the paper is as follows. In the next section, we introduce the basic notions and formally define k-approximations. To get started, in Section 3 we present the characterization of instance relationships in $\mathcal{EL}$ and show how this can be employed to compute k-approximations or (if it exists) the msc. In the subsequent section we extend the results to $\mathcal{EL}_\neg$, and finally deal with $\mathcal{ALE}$ in Section 5. The paper concludes with some remarks on future work. Due to space limitations, we refer to [14] for all technical details and complete proofs.

2 Preliminaries

Concept descriptions are inductively defined with the help of a set of *constructors*, starting with disjoint sets N_C of *concept names* and N_R of *role names*. In this work, we consider concept descriptions built from the constructors shown in Table 1, where $r \in N_R$ denotes a role name, $P \in N_C$ a concept name, and C, D concept descriptions. The concept descriptions in the DLs $\mathcal{EL}$, $\mathcal{EL}_\neg$, and $\mathcal{ALE}$ are built using certain subsets of these constructors, as shown in the last three columns of Table 1.

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 (disjoint from $N_C \cup N_R$), 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.

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 objects and the interpretation function $\cdot^\mathcal{I}$ maps each concept name $P \in N_C$ to a set $P^I \subseteq \Delta$, each role name $r \in N_R$ to a binary relation $r^I \subseteq \Delta \times \Delta$, and each individual $b \in N_I$ to an element $a^\mathcal{I} \in \Delta$ such that $a \neq b$ implies $a^\mathcal{I} \neq b^\mathcal{I}$ (*unique name assumption*). The extension of $\cdot^\mathcal{I}$ to arbitrary concept descriptions is inductively defined, as shown in the third column of Table 1. An interpretation $\mathcal{I}$ is a *model* of an ABox $\mathcal{A}$ 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 $a : C \in \mathcal{A}$.

The most important traditional inference services provided by DL-systems are computing the subsumption hierarchy and instance relationships. The concept description C is *subsumed* by the concept description D ($C \sqsubseteq D$) iff $C^I \subseteq D^I$ for all interpretations $\mathcal{I}$; C and D are *equivalent* ($C \equiv D$) iff they subsume each other. An individual $a \in N_I$ 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 $\mathcal{A}$.

In this paper, we are interested in the computation of most specific concepts and their approximation via concept descriptions of limited depth. The *depth* $\mathsf{depth}(C)$ of a concept description C is defined as the maximum of nested quantifiers in C. We also need to introduce least common subsumers formally.

Definition 1 (msc, k-approximation, lcs). *Let $\mathcal{A}$ be an $\mathcal{L}$-ABox, a an individual in $\mathcal{A}$, $C, C_1, \ldots, C_n$ $\mathcal{L}$-concept descriptions, and $k \in \mathbf{N}$. Then,*

- *C is the* most specific concept *(msc) of a w.r.t. $\mathcal{A}$ ($\mathsf{msc}_{\mathcal{A}}(a)$) iff $a \in_{\mathcal{A}} C$, and for all $\mathcal{L}$-concept descriptions C', $a \in_{\mathcal{A}} C'$ implies $C \sqsubseteq C'$;*
- *C is the* k-approximation *of a w.r.t. $\mathcal{A}$ ($\mathsf{msc}_{k,\mathcal{A}}(a)$) iff $a \in_{\mathcal{A}} C$, $\mathsf{depth}(C) \leq k$, and for all $\mathcal{L}$-concept descriptions C', $a \in_{\mathcal{A}} C'$ and $\mathsf{depth}(C') \leq k$ imply $C \sqsubseteq C'$;*
- *C is the* least common subsumer *(lcs) of $C_1, \ldots, C_n$ ($\mathsf{lcs}(C_1, \ldots, C_n)$) iff $C_i \sqsubseteq C$ for all $i = 1, \ldots, n$, and for all $\mathcal{L}$-concept descriptions C', $C_i \sqsubseteq C'$ for all $i = 1, \ldots, n$ implies $C \sqsubseteq C'$.*

Note that by definition, most specific concepts, k-approximations, and least common subsumers are uniquely determined up to equivalence (if they exist).

The following example shows that in DLs allowing for existential restrictions the msc of an ABox-individual b need not exist.

Example 1. Let $\mathcal{L}$ be one of the DLs $\mathcal{EL}$, $\mathcal{EL}_\neg$, or $\mathcal{ALE}$. Consider the $\mathcal{L}$-ABox $\mathcal{A} = \{(b, b) : r\}$. It is easy to see that, for each $n \geq 0$, b is an instance of the $\mathcal{L}$-concept description

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

The msc of b can be written as the infinite conjunction $\bigsqcap_{n \geq 0} C_n$, which, however, cannot be expressed by a (finite) $\mathcal{L}$-concept description. As we will see, the k-approximation of b is the $\mathcal{L}$-concept description $\bigsqcap_{0 \leq n \leq k} C_n$.

3 Most specific concepts in $\mathcal{EL}$

In the following subsection, we introduce the characterization of instance relationships in $\mathcal{EL}$, which yields the basis for the algorithm computing k-approximations (Section 3.2). All results presented in this section are rather straightforward. However, they prepare the ground for the more involved technical problems one encounters for $\mathcal{EL}_\neg$.

3.1 Characterizing instance in $\mathcal{EL}$

In order to characterize instance relationships, we need to introduce description graphs (representing ABoxes) and description trees (representing concept descriptions). For $\mathcal{EL}$, an *$\mathcal{EL}$-description graph* is a labeled graph of the form $\mathcal{G} = (V, E, \ell)$ whose edges $vrw \in E$ are labeled with role names $r \in N_R$ and whose nodes $v \in V$ are labeled with sets $\ell(v)$ of concept names from N_C. The empty label corresponds to the top-concept. An *$\mathcal{EL}$-description tree* is of the form $\mathcal{G} = (V, E, v_0, \ell)$, where (V, E, ℓ) is an $\mathcal{EL}$-description graph which is a tree with root $v_0 \in V$.

$\mathcal{EL}$-concept descriptions can be turned into $\mathcal{EL}$-description trees and vice versa [3]: Every $\mathcal{EL}$-concept description C can be written (modulo equivalence) as $C \equiv P_1 \sqcap \ldots \sqcap P_n \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_m.C_m$ with $P_i \in N_C \cup \{\top\}$. Such a concept description is recursively translated into an $\mathcal{EL}$-description tree $\mathcal{G}(C) = (V, E, v_0, \ell)$ as follows: The set of all concept names P_i occurring on the top-level of C yields the label $\ell(v_0)$ of the root v_0, and each existential restriction $\exists r_i.C_i$ yields an r_i-successor that is the root of the tree corresponding to C_i. For example, the concept description

$$C := \exists s.(Q \sqcap \exists r.\top) \sqcap \exists r.(Q \sqcap \exists s.\top)$$

yields the description tree depicted on the left hand side of Figure 1.

Every $\mathcal{EL}$-description tree $\mathcal{G} = (V, E, v_0, \ell)$ is translated into an $\mathcal{EL}$-concept description $C_\mathcal{G}$ as follows: the concept names occurring in the label of v_0 yield the concept names in the top-level conjunction of $C_\mathcal{G}$, and each $v_0 r v \in E$ yields an existential restriction $\exists r.C$, where C is the $\mathcal{EL}$-concept description obtained by translating the subtree of $\mathcal{G}$ with root v. For a leaf $v \in V$, the empty label is translated into the top-concept.

Adapting the translation of $\mathcal{EL}$-concept descriptions into $\mathcal{EL}$-description trees, an $\mathcal{EL}$-ABox $\mathcal{A}$ is translated into an $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ as follows: Let $\mathsf{Ind}(\mathcal{A})$ denote the set of all 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, $C_a := \top$. Let $\mathcal{G}(C_a) = (V_a, E_a, a, \ell_a)$ denote the $\mathcal{EL}$-description tree obtained from C_a. (Note that the individual a is defined to be the root of $\mathcal{G}(C_a)$; in particular, $a \in V_a$.) W.l.o.g. let the sets V_a, $a \in \mathsf{Ind}(\mathcal{A})$ be pairwise disjoint. Then, $\mathcal{G}(\mathcal{A}) := (V, E, \ell)$ is defined by

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

As an example, consider the $\mathcal{EL}$-ABox

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

The corresponding $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ is depicted on the right hand side of Figure 1.

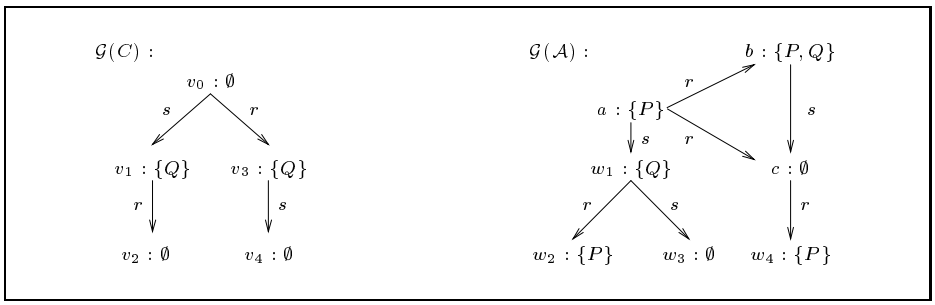

Fig. 1. The $\mathcal{EL}$-description tree of C and the $\mathcal{EL}$-description graph of $\mathcal{A}$.

Now, an instance relationship, $a \in_{\mathcal{A}} C$, in $\mathcal{EL}$ can be characterized in terms of a homomorphism from the description tree of C into the description graph of $\mathcal{A}$: a *homomorphism* from an $\mathcal{EL}$-description tree $\mathcal{G} = (V, E, v_0, \ell)$ into an $\mathcal{EL}$-description graph $\mathcal{H} = (V_H, E_H, \ell_H)$ is a mapping $\varphi : V \longrightarrow V_H$ such that (1) $\ell(v) \subseteq \ell_H(\varphi(v))$ for all $v \in V$, and (2) $\varphi(v) r \varphi(w) \in E_H$ for all $vrw \in E$.

Theorem 1. *[14] 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. Let $\mathcal{G}(\mathcal{A})$ denote the $\mathcal{EL}$-description graph of $\mathcal{A}$ and $\mathcal{G}(C)$ the $\mathcal{EL}$-description tree of C. Then, $a \in_{\mathcal{A}} C$ iff there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ such that $\varphi(v_0) = a$, where v_0 is the root of $\mathcal{G}(C)$.*

In our example, a is an instance of C, since mapping v_0 on a, v_i on w_i, $i = 1, 2$, and v_3 on b and v_4 on c yields a homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$.

Theorem 1 is a special case of the characterization of subsumption between simple conceptual graphs [6], and of the characterization of containment of conjunctive queries [1]. In these more general settings, testing for the existence of homomorphisms is an NP-complete problem. In the restricted case of testing homomorphisms mapping trees into graphs, the problem is polynomial [8]. Thus, as corollary of Theorem 1, we obtain the following complexity result.

Corollary 1. *The instance problem for $\mathcal{EL}$ can be decided in polynomial time.*

Theorem 1 generalizes the following characterization of subsumption in $\mathcal{EL}$ introduced in [3]. This characterization uses homomorphisms between description trees, which are defined just as homomorphisms from description trees into description graphs, but where we additionally require to map roots onto roots.

Theorem 2. *[3] Let C, D be $\mathcal{EL}$-concept descriptions, and let $\mathcal{G}(C)$ and $\mathcal{G}(D)$ be the corresponding description trees. Then, $C \sqsubseteq D$ iff there exists a homomorphism from $\mathcal{G}(D)$ into $\mathcal{G}(C)$.*

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

In the sequel, we assume $\mathcal{A}$ to be an $\mathcal{EL}$-ABox, a an individual occurring in $\mathcal{A}$, and k a non-negative integer. Roughly, our algorithm computing $\mathsf{msc}_{k,\mathcal{A}}(a)$

works as follows: First, the description graph $\mathcal{G}(\mathcal{A})$ is unraveled into a tree with root a. This tree, denoted $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$, has a finite branching factor, but possibly infinitely long paths. Pruning all paths to length k yields an $\mathcal{EL}$-description tree $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ of depth $\leq k$. Using Theorem 1 and Theorem 2, one can show that the $\mathcal{EL}$-concept description $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is equivalent to $\mathsf{msc}_{k,\mathcal{A}}(a)$. As an immediate consequence, in case $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is finite, $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ yields the msc of a. In what follows, we define $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ and $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$, and prove correctness of our algorithm.

First, we need some notations: For an $\mathcal{EL}$-description graph $\mathcal{G} = (V, E, \ell)$, $p = v_0 r_1 v_1 r_2 \dots r_n v_n$ is a *path from* v_0 *to* v_n of *length* $|p| = n$ and with *label* $r_1 \cdots r_n$, if $v_{i-1} r_i v_i \in E$ for all $1 \leq i \leq n$. The empty path ($n = 0$) is allowed, in which case the label of p is the empty word ε. The node v_n is an $r_1 \cdots r_n$-*successor of* v_0. Every node is an ε-successor of itself. A node v is *reachable* from v_0, if there exists a path from v_0 to v. The graph $\mathcal{G}$ is *cyclic*, if there exists a non-empty path from a node in $\mathcal{G}$ to itself.

Definition 2. *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*

- $V^T := \{p \mid p$ *is a path from a to some node in* $\mathcal{G}\}$,
- $E^T := \{prq \mid p, q \in V^T$ *and* $q = prw$ *for some* $r \in N_R$ *and* $w \in V\}$,
- $\ell^T(p) := \ell(v)$ *if p is a path to v.*

For $k \in \mathbf{N}$, 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*
- $\ell_k^T(p) := \ell^T(p)$ *for* $p \in V_k^T$.

Now, we can show the main theorem of this section.

Theorem 3. *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbf{N}$. Then, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is the k-approximation of a w.r.t. $\mathcal{A}$. If, starting from a, no cyclic path in $\mathcal{A}$ can be reached (i.e., $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is finite), then $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ is the msc of a w.r.t. $\mathcal{A}$; otherwise no msc exists.*

Proof sketch. Obviously, there exists a homomorphism from $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$, a tree isomorphic to $\mathcal{G}(C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))})$, into $\mathcal{G}(\mathcal{A})$ with a mapped on a. By Theorem 1, this implies $a \in_{\mathcal{A}} C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$.

Let C be an $\mathcal{EL}$-concept description with $a \in_{\mathcal{A}} C$ and $\mathsf{depth}(C) \leq k$. Theorem 1 implies that there exists a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$. Given φ, it is easy to construct a homomorphism from $\mathcal{G}(C)$ into $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$. Thus, with Theorem 2, we conclude $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))} \sqsubseteq C$. Altogether, this shows that $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is a k-approximation as claimed.

Now, assume that, starting from a, a cycle can be reached in $\mathcal{A}$, that is, $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is infinite. Then, we have a decreasing chain $C_0 \sqsupset C_1 \sqsupset \cdots$ of k-approximations C_k ($\equiv C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$) with increasing depth k, $k \geq 0$. From Theorem 2, we conclude that there does not exist an $\mathcal{EL}$-concept description

subsumed by all of these k-approximations (since such a concept description only has a fixed and finite depth). Thus, a cannot have an msc.

Conversely, if $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ is finite, say with depth k, from the observation that all k'-approximations, for $k' \geq k$, are equivalent, it immediately follows that $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))}$ is the msc of a. □

Obviously, there exists a deterministic algorithm computing the k-approximation (i.e., $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$) in time $\mathcal{O}(|\mathcal{A}|^k)$. The *size* $|\mathcal{A}|$ of $\mathcal{A}$ 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 occurrences of concept names, role names, and constructors in C. Similarly, one obtains an exponential complexity upper bound for computing the msc (if it exists).

Corollary 2. *For an $\mathcal{EL}$-ABox $\mathcal{A}$, an individual $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbb{N}$, the k-approximation of a w.r.t. $\mathcal{A}$ always exists and can be computed in time $\mathcal{O}(|\mathcal{A}|^k)$.*

The msc of a exists iff starting from a no cycle can be reached in $\mathcal{A}$. The existence of the msc can be decided in polynomial time, and if the msc exists, it can be computed in time exponential in the size of $\mathcal{A}$.

In the remainder of this section, we prove that the exponential upper bounds are tight. To this end, we show examples demonstrating that k-approximations and the msc may grow exponentially.

Example 2. Let $\mathcal{A} = \{(a,a) : r, (a,a) : s\}$. The $\mathcal{EL}$-description graph $\mathcal{G}(\mathcal{A})$ as well as the $\mathcal{EL}$-description trees $\mathcal{T}_1(a, \mathcal{G}(\mathcal{A}))$ and $\mathcal{T}_2(a, \mathcal{G}(\mathcal{A}))$ are depicted in Figure 2. It is easy to see that, for $k \geq 1$, $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ yields a full binary tree of depth k where

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

By Theorem 3, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is the k-approximation of a w.r.t. $\mathcal{A}$. The size of $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ is $|\mathcal{A}|^k$. Moreover, it is not hard to see 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 following example illustrates that, if it exists, also the msc can be of exponential size.

Example 3. 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 3, $C_{\mathcal{T}(a_1,\mathcal{A}_n)}$ 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 2. As before we obtain that

- $C_{\mathcal{T}(a_1,\mathcal{G}(\mathcal{A}))}$ is of size exponential in $|\mathcal{A}_n|$; and
- there does not exist an $\mathcal{EL}$-concept description C equivalent to but smaller than $C_{\mathcal{T}(a_1,\mathcal{G}(\mathcal{A}))}$.

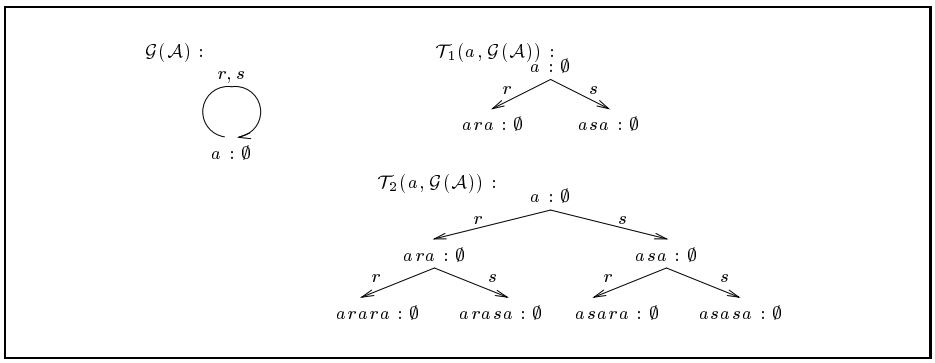

Fig. 2. The $\mathcal{EL}$-description graph and the $\mathcal{EL}$-description trees from Example 2.

Summarizing, we obtain the following lower bounds.

Proposition 1. *Let $\mathcal{A}$ be an $\mathcal{EL}$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbf{N}$.*

- *The size of $\mathsf{msc}_{\mathcal{A},k}(a)$ may grow with $|\mathcal{A}|^k$.*
- *If it exists, the size of $\mathsf{msc}_{\mathcal{A}}(a)$ may grow exponentially in $|\mathcal{A}|$.*

4 Most specific concepts in $\mathcal{EL}_\neg$

Our goal is to obtain a characterization of the (k-approximation of the) msc in $\mathcal{EL}_\neg$ analogously to the one given in Theorem 3 for $\mathcal{EL}$. To achieve this goal, first the notions of description graph and of description tree are extended from $\mathcal{EL}$ to $\mathcal{EL}_\neg$ by allowing for subsets of $N_C \cup \{\neg P \mid P \in N_C\} \cup \{\bot\}$ as node labels. Just as for $\mathcal{EL}$, there exists a 1–1 correspondence between $\mathcal{EL}_\neg$-concept descriptions and $\mathcal{EL}_\neg$-description trees, and an $\mathcal{EL}_\neg$-ABox $\mathcal{A}$ is translated into an $\mathcal{EL}_\neg$-description graph $\mathcal{G}(\mathcal{A})$ as described for $\mathcal{EL}$-ABoxes. The notion of a homomorphism also remains unchanged for $\mathcal{EL}_\neg$, and the characterization of subsumption extends to $\mathcal{EL}_\neg$ by just considering inconsistent $\mathcal{EL}_\neg$-concept descriptions as a special case: $C \sqsubseteq D$ iff $C \equiv \bot$ or there exists a homomorphism φ from $\mathcal{G}(D)$ into $\mathcal{G}(C)$.

Second, we have to cope with inconsistent $\mathcal{EL}_\neg$-ABoxes as a special case: for an inconsistent ABox $\mathcal{A}$, $a \in_{\mathcal{A}} C$ is valid for all concept descriptions C, and hence, $\mathsf{msc}_{\mathcal{A}}(a) \equiv \bot$. However, extending Theorem 1 with this special case does not yield a sound and complete characterization of instance relationships for $\mathcal{EL}_\neg$. If this was the case, we would get that the instance problem for $\mathcal{EL}_\neg$ is in P, in contradiction to complexity results shown in [16], which imply that the instance problem for $\mathcal{EL}_\neg$ is coNP-hard.

The following example is an abstract version of an example given in [16]; it illustrates incompleteness of a naïve extension of Theorem 1 from $\mathcal{EL}$ to $\mathcal{EL}_\neg$.

Example 4. 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 3. Obviously, there does not exist

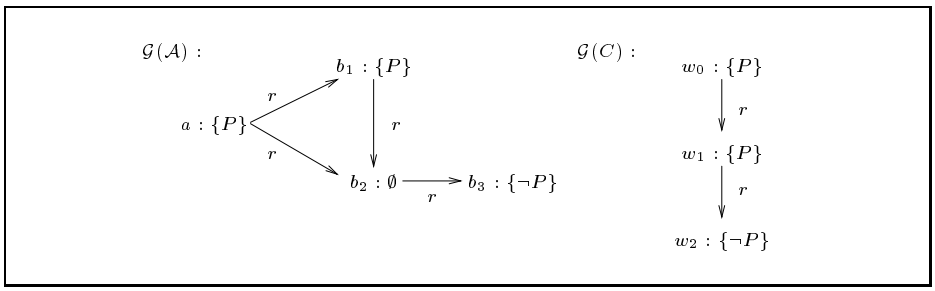

Fig. 3. The $\mathcal{EL}_\neg$-description graph and the $\mathcal{EL}_\neg$-description tree from Example 4.

a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = a$, because neither $P \in \ell(b_2)$ nor $\neg P \in \ell(b_2)$. 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 in fact, $a^\mathcal{I} \in C^\mathcal{I}$. Thus, 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 $\varphi(w_0) = a$.

In the following section, we give a sound and complete characterization of instance relationships in $\mathcal{EL}_\neg$, which again yields the basis for the characterization of k-approximations given in Section 4.2.

4.1 Characterizing instance in $\mathcal{EL}_\neg$

The reason for the problem illustrated in Example 4 is that, in general, for the individuals in the ABox it is not always fixed whether they are instances of a given concept name or not. Thus, in order to obtain a sound and complete characterization analogous to Theorem 1, instead of $\mathcal{G}(\mathcal{A})$, one has to consider all so-called atomic completions of $\mathcal{G}(\mathcal{A})$.

Definition 3 (Atomic 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 an* atomic completion *of $\mathcal{G}$ if, for all $v \in V$,*

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

Note that by definition, all labels of nodes in completions do not contain a conflict, i.e., the nodes are not labeled with a concept name and its negation. In particular, if $\mathcal{G}$ has a conflicting node, then $\mathcal{G}$ does not have a completion. It is easy to see that an $\mathcal{EL}_\neg$-ABox $\mathcal{A}$ is inconsistent iff $\mathcal{G}(\mathcal{A})$ contains a conflicting node. For this reason, in the following characterization of the instance relationship, we do not need to distinguish between consistent and inconsistent ABoxes.

Theorem 4. *[14] Let $\mathcal{A}$ be an $\mathcal{EL}_\neg$-ABox, $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$ the corresponding description graph, C an $\mathcal{EL}_\neg$-concept description, $\mathcal{G}(C) = (V_C, E_C, w_0, \ell_C)$ the corresponding description tree, and $a \in \mathsf{Ind}(\mathcal{A})$. Then, $a \in_\mathcal{A} C$ iff for each atomic 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 $\varphi(w_0) = a$.*

The problem of deciding whether there exists an atomic completion $\mathcal{G}(\mathcal{A})^*$ such that there exists no homomorphism from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})^*$ is in coNP. Adding the coNP-hardness result obtained from [16], this shows

Corollary 3. *The instance problem for $\mathcal{EL}_\neg$ is coNP-complete.*

4.2 Computing k-approximations in $\mathcal{EL}_\neg$

Not surprisingly, the algorithm computing the k-approximation/msc in $\mathcal{EL}$ does not yield the desired result for $\mathcal{EL}_\neg$. For instance, in Example 4, we would get $C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A}))} = P \sqcap \exists r.\exists r.(\neg P) \sqcap \exists r.(P \sqcap \exists r.\exists r.(\neg P))$. But as we will see, $\mathsf{msc}_\mathcal{A}(a) \equiv P \sqcap \exists r.(P \sqcap \exists r.\neg P) \sqcap \exists r.(P \sqcap \exists r.\exists r.\neg P)$, i.e., $\mathsf{msc}_\mathcal{A}(a) \sqsubset C_{\mathcal{T}(a,\mathcal{A})}$.

As in the extension of the characterization of instance relationships from $\mathcal{EL}$ to $\mathcal{EL}_\neg$, we have to take into account all atomic completions instead of the single description graph $\mathcal{G}(\mathcal{A})$. Intuitively, one has to compute the least concept description for which there exists a homomorphism into each atomic completion of $\mathcal{G}(\mathcal{A})$. In fact, this can be done by applying the lcs operation on the set of all concept descriptions $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^*)}$ obtained from the atomic completions $\mathcal{G}(\mathcal{A})^*$ of $\mathcal{G}(\mathcal{A})$.

Theorem 5. *Let $\mathcal{A}$ be an $\mathcal{EL}_\neg$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbf{N}$. If $\mathcal{A}$ is inconsistent, then $\mathsf{msc}_{k,\mathcal{A}}(a) \equiv \mathsf{msc}_\mathcal{A}(a) \equiv \bot$. Otherwise, let $\{\mathcal{G}(\mathcal{A})^1, \ldots, \mathcal{G}(\mathcal{A})^n\}$ be the set of all atomic completions of $\mathcal{G}(\mathcal{A})$.*

Then, $\mathsf{lcs}(C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^1)}, \ldots, C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^n)}) \equiv \mathsf{msc}_{k,\mathcal{A}}(a)$. If, starting from a, no cycle can be reached in $\mathcal{A}$, then $\mathsf{lcs}(C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^1)}, \ldots, C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^n)}) \equiv \mathsf{msc}_\mathcal{A}(a)$; otherwise the msc does not exist.

Proof sketch. Let $\mathcal{A}$ be a consistent $\mathcal{EL}_\neg$-ABox and $\mathcal{G}(\mathcal{A})^1, \ldots, \mathcal{G}(\mathcal{A})^n$ the atomic completions of $\mathcal{G}(\mathcal{A})$. By definition of $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)}$, there exists a homomorphism π_i from $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)}$ into $\mathcal{G}(\mathcal{A})^i$ for all $1 \leq i \leq n$. Let C_k denote the lcs of $\{C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^1)}, \ldots, C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^n)}\}$. The characterization of subsumption for $\mathcal{EL}_\neg$ yields homomorphisms φ_i from $\mathcal{G}(C_k)$ into $\mathcal{G}(C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)})$ for all $1 \leq i \leq n$. Now it is easy to see that $\pi_i \circ \varphi_i$ yields a homomorphism from $\mathcal{G}(C_k)$ into $\mathcal{G}(\mathcal{A})^i$, $1 \leq i \leq n$, each mapping the root of $\mathcal{G}(C_k)$ onto a. Hence, $a \in_\mathcal{A} C_k$.

Assume C' with $\mathsf{depth}(C') \leq k$ and $a \in_\mathcal{A} C'$. By Theorem 4, there exist homomorphisms ψ_i from $\mathcal{G}(C')$ into $\mathcal{G}(\mathcal{A})^i$ for all $1 \leq i \leq n$, each mapping the root of $\mathcal{G}(C')$ onto a. Since $\mathsf{depth}(C') \leq k$, these homomorphisms immediately yield homomorphisms ψ_i' from $\mathcal{G}(C')$ into $\mathcal{G}(C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)})$ for all $1 \leq i \leq n$. Now the characterization of subsumption yields $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A})^i)} \sqsubseteq C'$ for all $1 \leq i \leq n$, and hence $C_k \sqsubseteq C'$. Thus, $C_k \equiv \mathsf{msc}_{k,\mathcal{A}}(a)$.

Analogously, in case starting from a, no cycle can be reached in $\mathcal{A}$, we conclude $\mathsf{lcs}(C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^1)}, \ldots, C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^n)}) \equiv \mathsf{msc}_\mathcal{A}(a)$. Otherwise, with the same argument as in the proof of Theorem 3, it follows that the msc does not exist. □

In Example 4, we obtain two atomic completions, namely $\mathcal{G}(\mathcal{A})^1$ with $\ell^1(b_2) = \{P\}$, and $\mathcal{G}(\mathcal{A})^2$ with $\ell^2(b_2) = \{\neg P\}$. Now Theorem 5 implies $\mathsf{msc}_\mathcal{A}(a) \equiv \mathsf{lcs}(C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^1)}, C_{\mathcal{T}(a,\mathcal{G}(\mathcal{A})^2)})$, which is equivalent to

$$P \sqcap \exists r.(P \sqcap \exists r.\neg P) \sqcap \exists r.(P \sqcap \exists r.\exists r.\neg P).$$

The examples showing the exponential blow-up of the size of k-approximations and most specific concepts in $\mathcal{EL}$ can easily be adapted to $\mathcal{EL}_\neg$. However, we only have a double exponential upper bound (though we strongly conjecture that the size can again be bounded single-exponentially): the size of each tree (and the corresponding concept descriptions) obtained from an atomic completion is at most exponential, and the size of the lcs of a sequence of $\mathcal{EL}_\neg$-concept descriptions can grow exponentially in the size of the input descriptions [3].

Moreover, by an algorithm computing the lcs of the concept descriptions obtained from the atomic completions, the k-approximation (the msc) can be computed in double exponential time.

Corollary 4. *Let $\mathcal{A}$ be an $\mathcal{EL}_\neg$-ABox, $a \in \mathsf{Ind}(\mathcal{A})$, and $k \in \mathbf{N}$.*

- *The k-approximation of a always exists. It may be of size $|\mathcal{A}|^k$ and can be computed in double-exponential time.*
- *The msc of a exists iff $\mathcal{A}$ is inconsistent, or starting from a, no cycle can be reached in $\mathcal{A}$. If the msc exists, its size may grow exponentially in $|\mathcal{A}|$, and it can be computed in double-exponential time. The existence of the msc can be decided in polynomial time.*

5 Most specific concepts in $\mathcal{ALE}$

As already mentioned in the introduction, the characterization of instance relationships could not yet be extended from $\mathcal{EL}_\neg$ to $\mathcal{ALE}$. Since these structural characterizations were crucial for the algorithms computing the (k-approximation of the) msc in $\mathcal{EL}$ and $\mathcal{EL}_\neg$, no similar algorithms for $\mathcal{ALE}$ can be presented here. However, we show that

1. given that N_C and N_R are finite sets, the $\mathsf{msc}_{k,\mathcal{A}}(a)$ always exists and can effectively be computed (cf. Theorem 6);
2. the characterization of instance relationships in $\mathcal{EL}$ is also sound for $\mathcal{ALE}$ (cf. Lemma 1), which allows for approximating the k-approximation; and
3. we illustrate the main problems encountered in the structural characterization of instance relationships in $\mathcal{ALE}$ (cf. Example 5).

The first result is achieved by a rather generic argument. Given that the signature, i.e., the sets N_C and N_R, are fixed and finite, it is easy to see that also the set of $\mathcal{ALE}$-concept descriptions of depth $\leq k$ built using only names from $N_C \cup N_R$ is finite (up to equivalence) and can effectively be computed. Since the instance problem for $\mathcal{ALE}$ is known to be decidable [16], enumerating this set and retrieving the least concept description which has a as instance, obviously yields an algorithm computing $\mathsf{msc}_{k,\mathcal{A}}(a)$.

Theorem 6. *Let N_C and N_R be fixed and finite, and let $\mathcal{A}$ be an $\mathcal{ALE}$-ABox built over a set N_I of individuals and $N_C \cup N_R$. Then, for $k \in \mathbf{N}$ and $a \in \mathsf{Ind}(\mathcal{A})$, the k-approximation of a w.r.t. $\mathcal{A}$ always exists and can effectively be computed.*

Note that the above argument cannot be adapted to prove the existence of the msc for acyclic $\mathcal{ALE}$-ABoxes unless the size of the msc can be bounded appropriately. Finding such a bound remains an open problem.

The algorithm sketched above is obviously not applicable in real applications. Thus, in the remainder of this section, we focus on extending the improved algorithms obtained for $\mathcal{EL}$ and $\mathcal{EL}_\neg$ to $\mathcal{ALE}$.

5.1 Approximating the k-approximation in $\mathcal{ALE}$

We first have to extend the notions of description graph and of description tree from $\mathcal{EL}_\neg$ to $\mathcal{ALE}$: In order to cope with value restrictions occurring in $\mathcal{ALE}$-concept descriptions, we allow for two types of edges, namely those labeled with role names $r \in N_R$ (representing existential restrictions of the form $\exists r.C$) and those labeled with $\forall r$ (representing value restrictions of the form $\forall r.C$). Again, there is a 1–1 correspondence between $\mathcal{ALE}$-concept descriptions and $\mathcal{ALE}$-description trees, and an $\mathcal{ALE}$-ABox $\mathcal{A}$ is translated into an $\mathcal{ALE}$-description graph $\mathcal{G}(\mathcal{A})$ just as described for $\mathcal{EL}$-ABoxes. The notion of a homomorphism also extends to $\mathcal{ALE}$ in a natural way. 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 (1) and (2) on homomorphisms between $\mathcal{EL}$-description trees and $\mathcal{EL}$-description graphs, and additionally (3) $\varphi(v)\forall r\varphi(w) \in E$ for all $v\forall rw \in E_H$.

We are now ready to formalize soundness of the characterization of instance relationships for $\mathcal{ALE}$.

Lemma 1. *[14] Let $\mathcal{A}$ be an $\mathcal{ALE}$-ABox with $\mathcal{G}(\mathcal{A}) = (V, E, \ell)$, C an $\mathcal{ALE}$-concept description with $\mathcal{G}_C = (V_C, E_C, v_0, \ell_C)$, and $a \in \mathsf{Ind}(\mathcal{A})$.*

If there exists a homomorphism φ from $\mathcal{G}_C$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(v_0) = a$, then $a \in_{\mathcal{A}} C$.

As an immediate consequence of this lemma, we get $a \in_{\mathcal{A}} C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ for all $k \geq 0$, where the trees $\mathcal{T}(a, \mathcal{G}(\mathcal{A}))$ and $\mathcal{T}_k(a, \mathcal{G}(\mathcal{A}))$ are defined just as for $\mathcal{EL}$. This in turn yields $\mathsf{msc}_{k,\mathcal{A}}(a) \sqsubseteq C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))}$ and hence, an algorithm computing an approximation of the k-approximation for $\mathcal{ALE}$. In fact, such approximations already turned out to be quite usable in our process engineering application [4].

The following example now shows that the characterization is not complete for $\mathcal{ALE}$, and that, in general, $C_{\mathcal{T}_k(a,\mathcal{G}(\mathcal{A}))} \not\equiv \mathsf{msc}_{k,\mathcal{A}}(a)$. In particular, it demonstrates the difficulties one encounters in the presence of value restrictions.

Example 5. Consider the $\mathcal{ALE}$-ABox

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

and the $\mathcal{ALE}$-concept description $C = \exists r.(\forall s.P \sqcap \exists r.\exists s.\top)$; $\mathcal{G}(\mathcal{A})$ and $\mathcal{G}(C)$ are depicted in Figure 4. Note that $\mathcal{G}(\mathcal{A})$ is the unique atomic completion of itself (w.r.t. $N_C = \{P\}$).

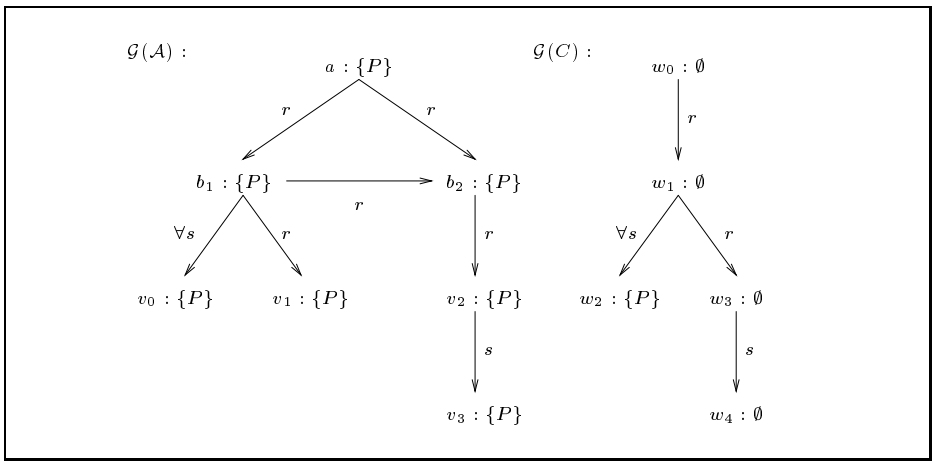

Fig. 4. The $\mathcal{ALE}$-description graph and the $\mathcal{ALE}$-description tree from Example 5.

It is easy to see that there does not exist a homomorphism φ from $\mathcal{G}(C)$ into $\mathcal{G}(\mathcal{A})$ with $\varphi(w_0) = a$. However, $a \in_{\mathcal{A}} C$: For each model $\mathcal{I}$ of $\mathcal{A}$, either $b_2^{\mathcal{I}}$ 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 $a^{\mathcal{I}}$. Thus, for each model $\mathcal{I}$ of $\mathcal{A}$, $a^{\mathcal{I}} \in C^{\mathcal{I}}$.

Moreover, for $k = 4$, $C_{\mathcal{T}_4(a,\mathcal{A})}$ is given by $P \sqcap \exists r.(P \sqcap \forall s.P \sqcap \exists r.P \sqcap \exists r.(P \sqcap \exists r.(P \sqcap \exists s.P))) \sqcap \exists r.(P \sqcap \exists r.(P \sqcap \exists s.P))$. It is easy to see that $C_{\mathcal{T}_4(a,\mathcal{A})} \not\sqsubseteq C$. Hence, $C_{\mathcal{T}_4(a,\mathcal{A})} \sqcap C \sqsubset C_{\mathcal{T}_4(a,\mathcal{A})}$, which implies $\mathsf{msc}_{4,\mathcal{A}}(a) \sqsubset C_{\mathcal{T}_4(a,\mathcal{A})}$.

Intuitively, the above example suggests that, in the definition of atomic completions, one should take into account not only (negated) concept names but also more complex concept descriptions. However, it is not clear whether an appropriate set of such concept descriptions can be obtained just from the ABox and how these concept descriptions need to be integrated in the completion in order to obtain a sound and complete structural characterization of instance relationships in $\mathcal{ALE}$.

6 Conclusion

Starting with the formal definition of the k-approximation of msc we showed that, for $\mathcal{ALE}$ and a finite signature (N_C, N_R), the k-approximation of the msc of an individual b always exists and can effectively be computed. For the sublanguages $\mathcal{EL}$ and $\mathcal{EL}_\neg$, we gave sound and complete characterizations of instance relationships that lead to practical algorithms. As a by-product, we obtained a characterization of the existence of the msc in $\mathcal{EL}$-/$\mathcal{EL}_\neg$-ABoxes, and showed that the msc can effectively be computed in case it exists.

First experiments with manually computed approximations of the msc in the process engineering application were quite encouraging [4]: used as inputs

for the lcs operation, i.e., the second step in the bottom-up construction of the knowledge base, they led to descriptions of building blocks the engineers could use to refine their knowledge base. In next steps, the run-time behavior and the quality of the output of the algorithms presented here is to be evaluated by a prototype implementation in the process engineering application.

References

1. S. Abiteboul, R. Hull, and V. Vianu. *Foundations of Databases*. Addison-Wesley, 1995.
2. 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 *Proc. of KI'98*, LNAI 1504, pp 129-140. Springer-Verlag, 1998.
3. F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In *Proc. of IJCAI'99*, pp 96-101. Morgan Kaufmann, 1999.
4. F. Baader and R. Molitor. Building and structuring description logic knowledge bases using least common subsumers and concept analysis. In *Proc. of ICCS2000*, LNAI 1867,pp 292-305. Springer-Verlag, 2000.
5. W.W. Cohen, A. Borgida, and H. Hirsh. Computing least common subsumers in description logics. In *Proc. of AAAI'92*, pp 754-760. MIT Press, 1992.
6. M. Chein and M. Mugnier. Conceptual graphs: Fundamental notions. *Revue d'Intelligence*. 6(4):365-406, 1992.
7. W.W. Cohen and H. Hirsh. Learning the CLASSIC description logic: Theoretical and experimental results. In *Proc. of KR'94*, pp 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. V. Haarslev and R. Möller. Expressive ABox reasoning with number restrictions, role hierarchies and transitively closed roles. In *Proc. of KR2000*, pp 273-284. Morgan Kaufmann, 2000.
10. I. Horrocks. Using an expressive description Logic: FaCT or fiction? In *Proc. of KR'98*, pp 636-647. Morgan Kaufmann, 1998.
11. R. Küsters. *Non-Standard Inference in Description Logics*. PhD thesis, RWTH Aachen, 2000. To appear as volume 2100 of the Springer Lecture Notes in Artificial Intelligence.
12. R. Küsters and A. Borgida. What's in an attribute? Consequences for the least common subsumer. *Journal of Artificial Intelligence Research*, 14: 167–203, 2001.
13. R. Küsters and R. Molitor. Computing least common subsumers in $\mathcal{ALEN}$. In *Proc. of IJCAI'01*. Morgan Kaufmann, 2001. To appear.
14. R. Küsters and R. Molitor. Computing most specific concepts in description logics with existential restrictions. Technical Report LTCS-00-05. See http://www-lti.informatik.rwth-aachen.de/Forschung/Reports.html.
15. B. Nebel. Reasoning and Revision in Hybrid Representation Systems. LNAI 422, Springer-Verlag, 1990.
16. 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.
17. L. von Wedel and W. Marquardt. ROME: A repository to support the integration of models over the lifecycle of model-based engineering processes. In *Proc. of ESCAPE-10*, 2000.