On Subsumption and Instance Problem in $\mathcal{ELH}$ w.r.t. General TBoxes

Sebastian Brandt*
Institut für Theoretische Informatik
TU Dresden, Germany
brandt@tcs.inf.tu-dresden.de

Abstract

Recently, it was shown for the DL $\mathcal{EL}$ that subsumption and instance problem w.r.t. cyclic terminologies can be decided in polynomial time. In this paper, we show that both problems remain tractable even when admitting general concept inclusion axioms and simple role inclusion axioms.

1 Motivation

In the area of DL based knowledge representation, the utility of *general* TBoxes, i.e., TBoxes that allow for general concept inclusion (GCI) axioms, is well known. For instance, in the context of the medical terminology Galen [18], GCIs are used especially for two purposes [16]:

- indicate the status of objects: instead of introducing several concepts for the same concept in different states, e.g., normal insulin secretion, abnormal but harmless insulin secretion, and pathological insulin secretion, only insulin secretion is defined while the status, i.e., normal, abnormal but harmless, and pathological, is implied by GCIs of the form $\ldots \sqsubseteq \exists \mathsf{has\_status}.\mathsf{pathological}$.
- to bridge levels of granularity and add implied meaning to concepts. A classical example [11] is to use a GCI like

$$\begin{aligned}&\mathsf{ulcer} \sqcap \exists \mathsf{has\_loc}.\mathsf{stomach}\\ &\quad\sqsubseteq \mathsf{ulcer} \sqcap \exists \mathsf{has\_loc}.(\mathsf{lining} \sqcap \exists \mathsf{is\_part\_of}.\mathsf{stomach})\end{aligned}$$

  to render the description of 'ulcer of stomach' more precisely to 'ulcer of lining of stomach' if it is known that 'ulcer of stomach' is specific of the lining of the stomach.

It has been argued that the use of GCIs facilitates the re-use of data in applications of different levels of detail while retaining all inferences obtained from the full description [18]. Hence, to examine reasoning w.r.t. general TBoxes has a strong practical motivation.

*Supported by the DFG under Grant BA 1122/4-3

Research on reasoning w.r.t. general TBoxes has mainly focused on very expressive DLs, reaching as far as, e.g., $\mathcal{ALCNR}$ [5] and $\mathcal{SHIQ}$ [12], in which deciding subsumption of concepts w.r.t. general TBoxes is EXPTIME hard. Fewer results exist on subsumption w.r.t. general terminologies in DLs below $\mathcal{ALC}$. In [9] the problem is shown to remain EXPTIME complete for a DL providing only conjunction, value restriction and existential restriction. The same holds for the small DL $\mathcal{AL}$ which allows for conjunction, value and unqualified existential restriction, and primitive negation [7]. Even for the simple DL $\mathcal{FL}_0$, which only allows for conjunction and value restriction, subsumption w.r.t. cyclic TBoxes with descriptive semantics is PSPACE hard [14], implying hardness for general TBoxes.

Recently, however, it was shown for the DL $\mathcal{EL}$ that subsumption and instance problem w.r.t. cyclic terminologies can be decided in polynomial time [3, 2]. In the present paper we show that even w.r.t. general $\mathcal{ELH}$-TBoxes, including GCIs and simple role inclusion axioms, subsumption and instance problem remain tractable. A surprising result given that DL systems usually employed for reasoning over general terminologies implement—highly optimized—EXPTIME algorithms [13, 10]. Similarly, Racer [10], the only practicable reasoner for ABox reasoning w.r.t. general TBoxes, uses an EXPTIME algorithm for the very expressive DL $\mathcal{ALCNH}_{R^+}$.

The paper is organized as follows. Basic definitions related to general $\mathcal{ELH}$ TBoxes are introduced in Section 2. In Sections 3 and 4 we show how to decide subsumption and instance problem, respectively, w.r.t. general $\mathcal{ELH}$-TBoxes in polynomial time. All details and full proofs of our results can be found in our technical report [4].

2 General TBoxes in $\mathcal{ELH}$

Concept descriptions are inductively defined with the help of a set of concept *constructors*, starting with a set N_{con} of *concept names* and a set N_{role} of *role names*. In this paper, we consider the DL $\mathcal{ELH}$ which provides the concept constructors top-concept ($\top$), conjunction ($C \sqcap D$), and existential restrictions ($\exists r.C$). As usual, $\mathcal{ELH}$ concept descriptions are interpreted w.r.t. a model-theoretic semantics, see [4] for details.

An $\mathcal{EL}$-terminology (called $\mathcal{EL}$-TBox) is a finite set $\mathcal{T}$ of axioms of the form $C \sqsubseteq D$ (called *GCI*) or $C \doteq D$ (called *definition* iff $C \in N_{\text{con}}$) or $r \sqsubseteq s$ (called *simple role inclusion axiom* (SRI)), where C and D are concept descriptions defined in $\mathcal{L}$ and $r, s \in N_{\text{role}}$. A concept name $A \in N_{\text{con}}$ is called *defined in* $\mathcal{T}$ iff $\mathcal{T}$ contains one or more axioms of the form $A \sqsubseteq D$ or $A \doteq D$. The *size* of $\mathcal{T}$ is defined as the sum of the sizes of all axioms in $\mathcal{T}$. Denote by $N_{\text{con}}^{\mathcal{T}}$ the set of all concept names occurring in $\mathcal{T}$ and by $N_{\text{role}}^{\mathcal{T}}$ the set of all role names occurring in $\mathcal{T}$. A TBox that may contain GCIs is called *general*. Denote by $\mathcal{ELH}$ the extension of $\mathcal{EL}$ by SRIs in TBoxes.

An interpretation $\mathcal{I}$ is a *model* of $\mathcal{T}$ iff for every GCI $C \sqsubseteq D \in \mathcal{T}$ it holds that $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$, for every definition $C \doteq D$ it holds that $C^{\mathcal{I}} = D^{\mathcal{I}}$, and for every SRI $r \sqsubseteq s$ it holds that $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$. A concept description C *subsumes* a concept description D w.r.t. $\mathcal{T}$ ($C \sqsubseteq_{\mathcal{T}} D$) iff $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ in every model $\mathcal{I}$ of $\mathcal{T}$. C and D are *equivalent* w.r.t. $\mathcal{T}$ ($C \equiv_{\mathcal{T}} D$) iff they subsume each other w.r.t. $\mathcal{T}$.

An $\mathcal{ELH}$-ABox is a finite set of assertions of the form $A(a)$ (called *concept assertion*) or $r(a,b)$ (called *role assertion*), where $A \in N_{\text{con}}$, $r \in N_{\text{role}}$, and a, b are *individual names* from a set N_{ind}. $\mathcal{I}$ is a model of a TBox $\mathcal{T}$ together with an ABox $\mathcal{A}$ iff $\mathcal{I}$ is a model of $\mathcal{T}$ and $a^{\mathcal{I}} \in \Delta^{\mathcal{I}}$ such that all assertions in $\mathcal{A}$ are satisfied, i.e., $a^{\mathcal{I}} \in A^{\mathcal{I}}$ for all $A(a) \in \mathcal{A}$ and $(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$ for all $r(a,b) \in \mathcal{A}$. An individual name a is an *instance of* C w.r.t. $\mathcal{T}$ ($\mathcal{A} \models_{\mathcal{T}} C(a)$) iff $a^{\mathcal{I}} \in A^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$ together with $\mathcal{A}$. Denote by $N_{\text{ind}}^{\mathcal{A}}$ the set of all individual names occurring in an ABox $\mathcal{A}$.

The above semantics for TBoxes and ABoxes is usually called *descriptive semantics* [15]. In case of an empty TBox, we write $C \sqsubseteq D$ instead of $C \sqsubseteq_{\emptyset} D$ and analogously $C \equiv D$ instead of $C \equiv_{\emptyset} D$.

Example 1 As an example of what can be expressed by an $\mathcal{ELH}$-TBox, consider the following TBox showing in an extremely simplified fashion a part of a medical terminology.

$$
\begin{aligned}
\mathsf{Pericardium} &\sqsubseteq \mathsf{Tissue} \sqcap \exists \mathsf{cont\_in}.\mathsf{Heart} \\
\mathsf{Pericarditis} &\sqsubseteq \mathsf{Inflammation} \\
&\quad \sqcap \exists \mathsf{has\_loc}.\mathsf{Pericardium} \\
\mathsf{Inflammation} &\sqsubseteq \mathsf{Disease} \sqcap \exists \mathsf{acts\_on}.\mathsf{Tissue} \\
\mathsf{Disease} \sqcap \exists \mathsf{has\_loc}.\exists \mathsf{comp\_of}.\mathsf{Heart} &\sqsubseteq \mathsf{Heartdisease} \\
&\quad \sqcap \exists \mathsf{is\_state}.\mathsf{NeedsTreatment} \\
\mathsf{cont\_in} &\sqsubseteq \mathsf{comp\_of}
\end{aligned}
$$

The TBox contains four GCIs and one SRI, stating, e.g., that Pericardium is tissue contained in the heart and that a diesease located in a component of the heart is a heart disease and requires treatment. Without going into detail, one can check that Pericarditis would be classified as a heart disease requiring treatment because, as stated in the TBox, Pericarditis is a disease located in the Pericardium contained in the heart, and everything contained in something is a component of it.[1]

3 Subsumption in $\mathcal{ELH}$ w.r.t. general TBoxes

We aim to show that subsumption of $\mathcal{ELH}$ concepts w.r.t. general TBoxes can be decided in polynomial time. A natural question is whether we may not simply utilize an existing decision procedure for a more expressive DL which might exhibit polynomial time complexity when applied to $\mathcal{ELH}$-TBoxes. Using the standard tableaux algorithm deciding consistency of general $\mathcal{ALC}$-TBoxes [1] as an example, one can show that this approach in general does not bear fruit, even for the sublanguage $\mathcal{EL}$, see [4].

Hence, new techniques are required exploiting the simpler structure of general $\mathcal{ELH}$-TBoxes better. The first step in our approach is to transform TBoxes into a normal form which limits the use of complex concept descriptions to the most basic cases.

[1]The example is only supposed to show the features of $\mathcal{ELH}$ and in no way claims to be adequate from a Medical KR point of view.

Definition 2 *(Normalized $\mathcal{ELH}$-TBox) Let $\mathcal{T}$ be an $\mathcal{ELH}$-TBox over N_{con} and N_{role}. $\mathcal{T}$ is* normalized *iff (i) $\mathcal{T}$ contains only GCIs and SRIs, and, (ii) all of the GCIs have one of the following forms:*

$$\begin{aligned} A &\sqsubseteq B \\ A_1 \sqcap A_2 &\sqsubseteq B \\ A &\sqsubseteq \exists r.B \\ \exists r.A &\sqsubseteq B, \end{aligned}$$

where A, A_1, A_2, B represent concept names from N_{con} or the top concept $\top$.

Such a normal form can be computed by exhaustively applying the following transformation rules.

Definition 3 *(Normalization rules) Let $\mathcal{T}$ be an $\mathcal{ELH}$-TBox over N_{con} and N_{role}. For every $\mathcal{ELH}$-concept description C, D, E over $N_{\text{role}} \cup \{\top\}$ and for every $r \in N_{\text{role}}$, the $\mathcal{ELH}$-normalization rules are defined modulo commutativity of conjunction ($\sqcap$) as follows:*

$$\begin{array}{llcl}
\textbf{NF1} & C \doteq D & \longrightarrow & \{C \sqsubseteq D,\ D \sqsubseteq C\} \\
\textbf{NF2} & \hat{C} \sqcap D \sqsubseteq E & \longrightarrow & \{\hat{C} \sqsubseteq A,\ A \sqcap D \sqsubseteq E\} \\
\textbf{NF3} & \exists r.\hat{C} \sqsubseteq D & \longrightarrow & \{\hat{C} \sqsubseteq A,\ \exists r.A \sqsubseteq D\} \\
\textbf{NF4} & C \sqsubseteq \exists r.\hat{D} & \longrightarrow & \{C \sqsubseteq \exists r.A,\ A \sqsubseteq \hat{D}\} \\
\textbf{NF5} & C \sqsubseteq D \sqcap E & \longrightarrow & \{C \sqsubseteq D,\ C \sqsubseteq E\}
\end{array}$$

where $\hat{C}, \hat{D}$ denote non-atomic concept descriptions and A denotes a new concept name from N_{con}. Applying a rule $G \longrightarrow \mathcal{S}$ to $\mathcal{T}$ changes $\mathcal{T}$ to $(\mathcal{T} \setminus \{G\}) \cup \mathcal{S}$. The normalized TBox norm$(\mathcal{T})$ is defined by exhaustively applying Rules **NF1** *to* **NF3** *and, after that, exhaustively applying Rules* **NF4** *and* **NF5**.

The size of $\mathcal{T}$ is increased only linearly by exhaustive application of Rule **NF1**. Since this rule never becomes applicable as a consequence of Rules **NF2** to **NF5**, we may restrict our attention to Rules **NF2** to **NF5**. A single application of one of the Rules **NF2** to **NF3** increases the size of $\mathcal{T}$ only by a constant, introducing a new concept name and splitting one GCI into two. Exhaustive application therefore produces an ontology of linear size in the size of $\mathcal{T}$.

After exhaustive application of Rules **NF1** to **NF3**, the left-hand side of every GCI is of constant size. Hence, applying Rules **NF4** and **NF5** exhaustively similarly yields an ontology of linear size in $\mathcal{T}$. Conseqently, the following lemma holds.

Lemma 4 *The normalized TBox norm$(\mathcal{T})$ can be computed in linear time in the size of $\mathcal{T}$. The resulting ontology is of linear size in the size of $\mathcal{T}$.*

ISR If $s \in S_i(r)$ and $s \sqsubseteq t \in \mathcal{T}$ and $t \notin S_{i+1}(r)$
then $S_{i+1}(r) := S_{i+1}(r) \cup \{t\}$

IS1 If $A_1 \in S_i(\alpha)$ and $A_1 \sqsubseteq B \in \mathcal{T}$ and $B \notin S_{i+1}(\alpha)$
then $S_{i+1}(\alpha) := S_{i+1}(\alpha) \cup \{B\}$

IS2 If $A_1, A_2 \in S_i(\alpha)$ and $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$
and $B \notin S_{i+1}(\alpha)$ then $S_{i+1}(\alpha) := S_{i+1}(\alpha) \cup \{B\}$

IS3 If $A_1 \in S_i(\alpha)$ and $A_1 \sqsubseteq \exists r.B \in \mathcal{T}$
and $B_1 \in S_i(B)$ and $s \in S_i(r)$ and $\exists s.B_1 \sqsubseteq C \in \mathcal{T}$
and $C \notin S_{i+1}(\alpha)$ then $S_{i+1}(A) := S_{i+1}(\alpha) \cup \{C\}$

Figure 1: Rules for implication sets

Note that applying Rule **NF5** before exhaustive application of the other rules may produce a terminology of quadratic size in the size of $\mathcal{T}$.

Our strategy is, for every concept name $A \in N_{\text{con}}^{\mathcal{T}}$ and $\top$, to compute a set of concept names $S_*(A)$ with the following property: whenever in some point x in a model of $\mathcal{T}$ the concept A holds then every concept in $S_*(A)$ necessarily also holds in x. Similarly, for every role r we want to represent by $S_*(r)$ the set of all roles included in r. The simple structure of GCIs in normalized TBoxes allows us to define such sets as follows. To simplify notation, let $N_{\text{con}}^{\mathcal{T},\top} := N_{\text{con}}^{\mathcal{T}} \cup \{\top\}$.

Definition 5 *(Implication set) Let $\mathcal{T}$ denote a normalized $\mathcal{ELH}$-TBox $\mathcal{T}$ over N_{con} and N_{role}. For every $A \in N_{\text{con}}^{\mathcal{T},\top}$ ($r \in N_{\text{role}}^{\mathcal{T}}$) and every $i \in \mathbb{N}$, the set $S_i(A)$ ($S_i(r)$) is defined inductively, starting by $S_0(A) := \{A, \top\}$ ($S_0(r) := \{r\}$). For every $i \geq 0$, $S_{i+1}(A)$ ($S_{i+1}(r)$) is obtained by extending $S_i(A)$ ($S_i(r)$) by exhaustive application of the extension rules shown in Figure 1, where $\alpha \in N_{\text{con}}^{\mathcal{T},\top}$. The* implication set *$S_*(A)$ of A is defined as the infinite union $S_*(A) := \bigcup_{i \geq 0} S_i(A)$. Analogously, define $S_*(r) := \bigcup_{i \geq 0} S_i(r)$.*

Note that the successor $S_{i+1}(A)$ of some $S_i(A)$ is generally not the result of only a *single* rule application. $S_{i+1}(A)$ is complete only if no more rules are applicable to any $S_i(B)$ or $S_i(r)$. Implication sets induce a reflexive and transitive but not symmetric relation on $N_{\text{con}}^{\mathcal{T},\top}$ and $N_{\text{role}}^{\mathcal{T}}$, since $B \in S_*(A)$ does not imply $A \in S_*(B)$. We have to show that the idea underlying implication sets is indeed correct. Hence, the occurrence of a concept name B in $S_*(A)$ implies that $A \sqsubseteq_{\mathcal{T}} B$ and vice versa.

Lemma 6 *For every normalized $\mathcal{ELH}$-TBox over N_{con} and N_{role}, (i) for every $r, s \in N_{\text{role}}^{\mathcal{T}}$, $s \in S_*(r)$ iff $r \sqsubseteq_{\mathcal{T}} s$, and (ii) for every $A, B \in N_{\text{con}}^{\mathcal{T},\top}$ it holds that $B \in S_*(A)$ iff $A \sqsubseteq_{\mathcal{T}} B$.*

We give a proof sketch, the full proof is shown in [4]. For Claim (i), obviously $r \sqsubseteq_{\mathcal{T}} s$ iff (r, s) is in the transitive closure induced by all $s' \sqsubseteq t' \in \mathcal{T}$. Exactly this closure is computed breadth-first by means of Rule **ISR**.

For the direction ($\Rightarrow$) of Claim (ii), assume $x \in A^{\mathcal{I}}$ for some model $\mathcal{I}$ of $\mathcal{T}$ and $B \in S_*(A)$. Proof by induction over the minimal n with $B \in S_n(A)$. For $n = 0$, $B \in$

$\{A, \top\}$, implying $x \in B^{\mathcal{I}}$. For $n > 0$, we distinguish the rule which caused the inclusion of B in the ith step. In each case the induction hypothesis for the precondition of Rule **IS1** to **IS3** implies the semantical consequence $x \in B^{\mathcal{I}}$. For instance, if B has been included in $S_n(A)$ as a result of Rule **IS3** then there exist concept names $A_1, A_2, A_3 \in N_{\text{con}}^{\mathcal{T},\top}$ such that, on the one hand, $A_1 \in S_{n-1}(A)$ and $G := A_1 \sqsubseteq \exists r.A_2 \in \mathcal{T}$, and on the other hand, $A_3 \in S_{n-1}(A_2)$ and $H := \exists s.A_3 \sqsubseteq B \in \mathcal{T}$ with $s \in S_{n-1}(r)$. By induction hypothesis, $r \sqsubseteq_{\mathcal{T}} s$, implying by G that $x \in (\exists r.A_2)^{\mathcal{I}}$. Since $A_3 \in S_{n-1}(A_2)$ the induction hypothesis implies $x \in A_1^{\mathcal{I}}$ and $x \in (\exists s.A_3)^I$, yielding by H that $x \in B^{\mathcal{I}}$.

The reverse direction ($\Leftarrow$) is more involved. We show that if $B \notin S_*(A)$ then there is a model $\mathcal{I}$ of $\mathcal{T}$ with a witness $x_A \in A^{\mathcal{I}} \setminus B^{\mathcal{I}}$. We construct a canonical model $\mathcal{I}$ for A starting from a single vertex $x_A \in A^{\mathcal{I}}$, iteratively applying generation rules which extend $\mathcal{I}$ so as to satisfy all GCIs in $\mathcal{T}$. As $\mathcal{T}$ is normalized, one rule for each type of GCI suffices. For instance, a GCI $A \sqsubseteq \exists r.B$ induces for $x \in A^{\mathcal{I}}$ the creation of an r-successor labeled B. After showing that the (possibly infinite) model thus constructed is in fact a model of A, we show by induction over the construction of $\mathcal{I}$ that the following property holds for every vertex x. If A is the first concept name to whose interpretation x was added and if also $x \in B^{\mathcal{I}}$ then $B \in S_*(A)$. Note that this holds in general only if A is the 'oldest' concept with $x \in A^{\mathcal{I}}$. The induction step exploits the fact that if a generation rule for $\mathcal{I}$ forces x into the extension of B then one of the Rules **IS1** to **IS3** includes B into some $S_m(A)$. For instance, in the most simple case, if $x \in B^{\mathcal{I}}$ because of a GCI $C \sqsubseteq B$ then at some point previous, $x \in C^{\mathcal{I}}$, implying $C \in S_*(A)$ by induction hypothesis, yielding $B \in S_*(A)$ by Rule **IS1**.

To show decidability in polynomial time it suffices to show that, (i) $\mathcal{T}$ can be normalized in polynomial time (see above), and, (ii) for all $A \in N_{\text{con}}^{\mathcal{T},\top}$ and $r \in N_{\text{role}}^{\mathcal{T}}$, the sets $S_*(A)$ and $S_*(r)$ can be computed in polynomial time in the size of $\mathcal{T}$. Every $S_{i+1}(A)$ and $S_{i+1}(r)$ depends only on sets with index i. Hence, once $S_{i+1}(A) = S_i(A)$ and $S_{i+1}(r) = S_i(r)$ holds for all A and r the complete implication sets are obtained. This happens after a polynomial number of steps, since $S_i(A) \subseteq N_{\text{con}}^{\mathcal{T}}$ and $S_i(r) \subseteq N_{\text{role}}^{\mathcal{T}}$. To compute $S_{i+1}(A)$ and $S_{i+1}(r)$ from the $S_i(B)$ and $S_i(s)$ costs only polynomial time in the size of $\mathcal{T}$.

Theorem 7 *Subsumption in $\mathcal{ELH}$ w.r.t. general TBoxes can be decided in polynomial time.*

4 The instance problem in $\mathcal{ELH}$ w.r.t. general TBoxes

We show that the instance problem in $\mathcal{ELH}$ w.r.t. general TBoxes can be decided in polynomial time. To this end, the approach to decide subsumption by means of implication sets for concept names presented in the previous section is extended to ABox individuals. For every individual name $a \in N_{\text{ind}}^{\mathcal{A}}$, we want to compute a set $S_*(a)$ of concept names with the following property: if $A \in S_*(a)$ then in every model $\mathcal{I}$ of $\mathcal{T}$ together with $\mathcal{A}$ the individual $a^{\mathcal{I}}$ is a witness of A (and vice versa). To extend the definition of implication sets in this way we generalize Rules **IS1** to **IS3** to individual names and introduce a new Rule **IS4** specifically for individual names.

IS4 If $r(a,b) \in \mathcal{A}$ and $B \in S_i(b)$ and $s \in S_i(r)$ and $\exists s.B \sqsubseteq C \in \mathcal{T}$ and $C \notin S_{i+1}(a)$ then $S_{i+1}(a) := S_{i+1}(a) \cup \{C\}$

Figure 2: Additional rule for implication sets (instance problem)

Definition 8 *(Implication set) Let $\mathcal{T}$ denote a normalized $\mathcal{ELH}$-TBox $\mathcal{T}$ over N_{con} and N_{role} and $\mathcal{A}$ an ABox over N_{ind}, $N_{\text{con}}^{\mathcal{T}}$ and $N_{\text{role}}^{\mathcal{T}}$. For every $r \in N_{\text{role}}^{\mathcal{T}}$, $A \in N_{\text{con}}^{\mathcal{T},\top}$, and $a \in N_{\text{ind}}^{\mathcal{A}}$ and for every $i \in \mathbb{N}$, the sets $S_i(r)$, $S_i(A)$, and $S_i(a)$ are defined inductively, starting by*

$$
\begin{aligned}
S_0(r) &:= \{r\} \\
S_0(A) &:= \{A, \top\} \\
S_0(a) &:= \{A \mid A(a) \in \mathcal{A}\} \cup \{\top\}.
\end{aligned}
$$

For every $i \geq 0$, $S_{i+1}(r)$, $S_{i+1}(A)$, and $S_{i+1}(a)$ are obtained by extending $S_i(r)$, $S_i(A)$, and $S_i(a)$, respectively, by exhaustive application of Rules **ISR** *to* **IS4** *shown in Figures 1 and 2, where $\alpha \in N_{\text{con}}^{\mathcal{T},\top} \cup N_{\text{ind}}^{\mathcal{A}}$. The* implication set *$S_*(r)$ of r is defined as the infinite union $S_*(r) := \bigcup_{i\geq 0} S_i(r)$. Analogously, define $S_*(A) := \bigcup_{i\geq 0} S_i(A)$ and $S_*(a) := \bigcup_{i\geq 0} S_i(a)$.*

Since the above definition extends Definition 5 without adding new rules for concept-implication sets $S_*(A)$, Lemma 6 still holds. The following lemma shows that the idea underlying individual-implication sets $S_*(a)$ is also correct in the sense that $A \in S_*(a)$ iff $\mathcal{A} \models_{\mathcal{T}} A(a)$. W.l.o.g. we assume that every individual name $a \in N_{\text{ind}}^{\mathcal{A}}$ has at most one concept assertion $A(a) \in \mathcal{A}$. For every a with $\{A_1(a), A_2(a)\} \subseteq \mathcal{A}$ this can be satisfied by (i) introducing new TBox definitions of the form $A_a \sqsubseteq A_1 \sqcap A_2$ and $A_1 \sqcap A_2 \sqsubseteq A_a$, where A_a is a new concept name, and, (ii) modifying $\mathcal{A}$ to $(\mathcal{A} \setminus \{A_1(a), A_2(a)\}) \cup \{A_a(a)\}$. Iterating this modification yields a normalized TBox $\mathcal{T}'$ of linear size in $\mathcal{T}$ with the required property.

Lemma 9 *Let $\mathcal{T}$ be a normalized $\mathcal{ELH}$-TBox over N_{con} and N_{role} and $\mathcal{A}$ an ABox over N_{ind}, $N_{\text{con}}^{\mathcal{T}}$ and $N_{\text{role}}^{\mathcal{T}}$. For every $A_0 \in N_{\text{con}}^{\mathcal{T}}$ and every $a_0 \in N_{\text{ind}}^{\mathcal{A}}$, $A_0 \in S_*(a_0)$ iff $\mathcal{A} \models_{\mathcal{T}} A_0(a_0)$.*

Similar to Lemma 6, proof direction ($\Rightarrow$) is shown by induction over the least n for which $A_0 \in S_n(a_0)$. For the more interesting reverse direction ($\Leftarrow$), we assume $A_0 \notin S_*(a_0)$ and construct a canonical model $\mathcal{I}$ of $\mathcal{T}$ together with $\mathcal{A}$ where $a^{\mathcal{I}} \notin A^{\mathcal{I}}$. See [4] for the full proof.

The proof of decidability in polynomial time is analogous to the case of subsumption: regarding computational complexity, the individual-implication sets $S_*(a)$ have the same properties as concept-implication sets. The new Rule **IS4** also does not increase the complexity of computing the sets $S_*(a)$ significantly.

Theorem 10 *The instance problem in $\mathcal{ELH}$ w.r.t. general TBoxes can be decided in polynomial time.*

5 Conclusion

We have seen how subsumption and instance problem in $\mathcal{ELH}$ w.r.t. general TBoxes can be decided in polynomial time. Moreover, the implication sets computed for one TBox $\mathcal{T}$ can be used to decide *all* subsumptions between defined (or primitive) concepts in $\mathcal{T}$. Hence, classifying $\mathcal{T}$ requires only a single computation of the implication sets for $\mathcal{T}$. The same holds for the instance problem, where a single computation of the relevant implication sets suffices to classify $\mathcal{T}$ and decide *all* instance problems w.r.t. defined (or primitive) concepts occurring in $\mathcal{T}$.

Since subsumption and instance problem remain tractable under the transition from cyclic to general $\mathcal{EL}$-TBoxes, the second natural question is how far the DL can be extended further preserving tractability. Obviously, adding value restrictions makes subsumption NP hard even for the empty TBox [8]. Moreover, it can be shown that adding one of the constructors number restriction, disjunction, or allsome [6] makes subsumption co-NP hard even without GCIs.

It is open, however, whether subsumption and instance problem w.r.t. general TBoxes remain tractable when extending $\mathcal{ELH}$ by inverse roles. Extending our subsumption algorithm by more expressive role constructors might lead the way to a more efficient reasoning algorithm for the representation language underlying the Galen [17] terminology, where inverse roles and complex role inclusion axioms can be expressed. While the polynomial upper bound would undoubtedly be exceeded, still a complexity better than EXPTIME might be feasible.

Acknowledgements

My thanks to Carsten Lutz for a multitude of useful remarks and ideas that have greatly influenced this work.

References

[1] F. Baader and U. Sattler. An overview of tableau algorithms for description logics. *Studia Logica*, 69:5–40, 2001.

[2] Franz Baader. The instance problem and the most specific concept in the description logic $\mathcal{EL}$ w.r.t. terminological cycles with descriptive semantics. In *Proceedings of the 26th Annual German Conference on Artificial Intelligence, KI 2003*, volume 2821 of *Lecture Notes in Artificial Intelligence*, pages 64–78, Hamburg, Germany, 2003. Springer-Verlag.

[3] Franz Baader. Terminological cycles in a description logic with existential restrictions. In Georg Gottlob and Toby Walsh, editors, *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 325–330. Morgan Kaufmann, 2003.

[4] S. Brandt. Subsumption and instance problem in $\mathcal{ELH}$ w.r.t. general tboxes. LTCS-Report LTCS-04-04, Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology, Germany, 2004. See http://lat.inf.tu-dresden.de/research/reports.html.

[5] M. Buchheit, F. M. Donini, and A. Schaerf. Decidable reasoning in terminological knowledge representation systems. *Journal of Artificial Intelligence Research*, 1:109–138, 1993.

[6] Robert Dionne, Eric Mays, and Frank J. Oles. The equivalence of model-theoretic and structural subsumption in description logics. In Ruzena Bajcsy, editor, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 710–716, San Mateo, California, 1993. Morgan Kaufmann.

[7] F.M. Donini. Complexity of reasoning. In Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors, *The Description Logic Handbook: Theory, Implementation, and Applications*, pages 96–136. Cambridge University Press, 2003.

[8] Francesco M. Donini, Maurizio Lenzerini, Daniele Nardi, Bernhard Hollunder, Werner Nutt, and Alberto Spaccamela. The complexity of existential quantification in concept languages. *Artificial Intelligence*, 53(2–3):309–327, 1992.

[9] Robert Givan, David A. McAllester, Carl Witty, and Dexter Kozen. Tarskian set constraints. *Information and Computation*, 174(2):105–131, 2002.

[10] Volker Haarslev and Ralf Möller. RACER system description. *Lecture Notes in Computer Science*, 2083:701–712, 2001.

[11] Ian Horrocks, Alan L. Rector, and Carole A. Goble. A description logic based schema for the classification of medical data. In *Knowledge Representation Meets Databases*, 1996.

[12] Ian Horrocks, Ulrike Sattler, and Stephan Tobies. Practical reasoning for expressive description logics. In Harald Ganzinger, David McAllester, and Andrei Voronkov, editors, *Proceedings of the 6th International Conference on Logic for Programming and Automated Reasoning (LPAR'99)*, number 1705 in Lecture Notes in Artificial Intelligence, pages 161–180. Springer-Verlag, September 1999.

[13] Ian R. Horrocks. Using an expressive description logic: FaCT or fiction? In Anthony G. Cohn, Lenhart Schubert, and Stuart C. Shapiro, editors, *KR'98: Principles of Knowledge Representation and Reasoning*, pages 636–645. Morgan Kaufmann, San Francisco, California, 1998.

[14] Yevgeny Kazakov and Hans De Nivelle. Subsumption of concepts in $\mathcal{FL}_0$ for (cyclic) terminologies with respect to descriptive semantics is pspace-complete. In *Proceedings of the 2003 International Workshop on Description Logics (DL2003)*, CEUR-WS, 2003.

[15] B. Nebel. Terminological cycles: Semantics and computational properties. In J. F. Sowa, editor, *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, pages 331–361. Morgan Kaufmann Publishers, San Mateo (CA), USA, 1991.

[16] A. Rector. Medical informatics. In Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors, *The Description Logic Handbook: Theory, Implementation, and Applications*, pages 406–426. Cambridge University Press, 2003.

[17] A. Rector, S. Bechhofer, C. A. Goble, I. Horrocks, W. A. Nowlan, and W. D. Solomon. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*, 9:139–171, 1997.

[18] A. Rector, W. Nowlan, and A. Glowinski. Goals for concept representation in the GALEN project. In *Proceedings of the 17th annual Symposium on Computer Applications in Medical Care, Washington, USA, SCAMC*, pages 414–418, 1993.