Reasoning in $\mathcal{ELH}$ w.r.t. General Concept Inclusion Axioms

Sebastian Brandt
Theoretical Computer Science
TU Dresden
brandt@tcs.inf.tu-dresden.de

Abstract

In the area of Description Logic (DL) based knowledge representation, research on reasoning w.r.t. general terminologies has mainly focused on very expressive DLs. Recently, though, it was shown for the DL $\mathcal{EL}$, providing only the constructors conjunction and existential restriction, that the subsumption problem w.r.t. cyclic terminologies can be decided in polynomial time, a surprisingly low upper bound. In this paper, we show that even admitting general concept inclusion (GCI) axioms and role hierarchies in $\mathcal{EL}$ terminologies preserves the polynomial time upper bound for subsumption. We also show that subsumption becomes co-NP hard when adding one of the constructors number restriction, disjunction, and 'allsome', an operator used in the DL K-REP. An interesting implication of the first result is that reasoning over the widely used medical terminology SNOMED is possible in polynomial time.

Contents

1 Motivation

In the area of Description Logic (DL) based knowledge representation, intensional knowledge of a problem domain is represented in the form of a terminology (TBox) which declares general properties of concepts relevant to the domain [19]. In its most basic form, a TBox contains concept *definitions* of the form $A \doteq C$ which define a concept *name* A by a concept *description* C. Concept descriptions are terms built from primitive concepts by means of language constructors provided by the DL. The meaning of A w.r.t. the TBox is defined by interpreting the TBox w.r.t. a model-theoretic *semantics*, which allows formally well-defined reasoning over the terminology.

In addition, *general* TBoxes can contain universally true implications, so-called *general concept inclusion (GCI)* axioms of the form $C \sqsubseteq D$, where both C and D are arbitrary concept descriptions. A model respects a GCI $C \sqsubseteq D$ iff the extension of C is a subset of the extension of D. Hence, D is implied whenever C holds.

From an application point of view, the utility of general TBoxes for DL knowledge bases has long been observed. For instance, in the context of the medical terminology Galen [24], GCIs are used especially for two purposes [22]:

- 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 [14] 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 [24]. Hence, to examine reasoning w.r.t. general TBoxes has a strong practical motivation.

There is also a strong motivation to consider the DL $\mathcal{EL}$, providing only the constructors conjunction and existential restriction. The widely used medical terminology SNOMED [8] corresponds to an $\mathcal{EL}$ TBox [25]. The representation language underlying the medical terminology GALEN [24] in which GCIs are used extensively, similarly can be represented by a general $\mathcal{EL}$ TBoxe, requiring additional constructs for roles, though.

Research on reasoning w.r.t. general TBoxes has been mainly focused on very expressive DLs, reaching as far as, e.g., $\mathcal{ALCNR}$ [7] and $\mathcal{SHIQ}$ [15], in which deciding subsumption of concepts w.r.t. general TBoxes is EXPTIME hard. Fewer results exist on subsumption w.r.t. general terminologies DLs below $\mathcal{ALC}$. In [12] 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 [10]. 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 [17], implying hardness for general TBoxes.

Recently, however, it was shown for the DL $\mathcal{EL}$ that the subsumption problem w.r.t. cyclic terminologies can be decided in polynomial time [6]. Given the practical utility of general TBoxes on the one hand and this surprisingly low upper bound on the other, the present paper aims to explore how far the polynomial time bound reaches when extending cyclic $\mathcal{EL}$-TBoxes further. We show that admitting both GCIs and simple role inclusion axioms at the same time preserves the upper bound for subsumption. In contrast, by extending $\mathcal{EL}$ by one of the constructors number restriction, disjunction, and allsome, subsumption is shown to be co-NP hard.

The paper is organized as follows. Section 2 introduces basic notions essential to study the DLs under consideration. In Section 3 we present a polynomial time algorithm to decide subsumption in $\mathcal{ELH}$ w.r.t. general TBoxes and simple role inclusion axioms. Section 4 is dedicated to the co-NP hard extensions of $\mathcal{EL}$.

2 Description Logics

Syntax	Semantics	$\mathcal{EL}$	$\mathcal{ELU}$	$\mathcal{ELN}$	$\mathcal{EL}_{\forall\exists}$	$\mathcal{L}_{\forall\exists}$
$\top$	$\Delta^{\mathcal{I}}$	x	x	x	x	x
$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$	x	x	x	x	x
$C \sqcup D$	$C^{\mathcal{I}} \cup D^{\mathcal{I}}$		x			
$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y\colon (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$	x	x	x	x	
$\forall r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \forall y\colon (x,y) \in r^{\mathcal{I}} \Rightarrow y \in C^{\mathcal{I}}\}$					
$\forall\exists r.C$	$\forall r.C \sqcap \exists r.C$				x	x
$(\leq n\, r)$, $n \in \mathbb{N}$	$\{x \in \Delta^{\mathcal{I}} \mid \#\{y \mid (x,y) \in r^{\mathcal{I}}\} \leq n\}$			x		
$(\geq n\, r)$, $n \in \mathbb{N}$	$\{x \in \Delta^{\mathcal{I}} \mid \#\{y \mid (x,y) \in r^{\mathcal{I}}\} \geq n\}$			x		

Table 1: Syntax and semantics of concept descriptions.

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 concept descriptions built from the constructors shown in Table 1. All concept descriptions under consideration provide the constructors top-concept ($\top$) and conjunction ($C \sqcap D$) but otherwise differ from one another. Our point of departure will be the DL $\mathcal{EL}$ which also allows for existential restrictions ($\exists r.C$). The DL $\mathcal{ELU}$ extends $\mathcal{EL}$ by disjunction ($\sqcup$) while $\mathcal{ELN}$ extends $\mathcal{EL}$ by number restrictions ($\geq n\, r$) and ($\leq n\, r$). The DL $\mathcal{EL}_{\forall\exists}$ extends $\mathcal{EL}$ by the constructor allsome ($\forall\exists r.C$). The DL $\mathcal{L}_{\forall\exists}$ is obtained by removing existential restrictions from $\mathcal{EL}_{\forall\exists}$. (see Table 1).

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

For a given the DL $\mathcal{L}$, an $\mathcal{L}$-terminology (called $\mathcal{L}$-TBox) is a finite set $\mathcal{T}$ of axioms of the form $C \sqsubseteq D$ (called *GCI*) or $C \doteq D$ (called *definition*) 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 contains GCIs is called *general*. Denote by $\mathcal{ELH}$ the DL $\mathcal{EL}$ admitting 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 is *satisfiable* w.r.t. $\mathcal{T}$ iff there exists a model $\mathcal{I}$ such that $C^{\mathcal{I}} \neq \emptyset$. 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. $\mathcal{T}$. This semantics for TBoxes is usually called *descriptive semantics* [20]. 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 with 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]

[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.

3 Reasoning in $\mathcal{ELH}$ with GCIs

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 [4] as an example, one can show that this approach in general does not bear fruit, eveb for the sublanguage $\mathcal{EL}$.

In order to decide subsumption $C \sqsubseteq^?_{\mathcal{T}} D$ w.r.t. an $\mathcal{EL}$-TBox, an intuitive decision procedure to choose would be the $\mathcal{ALC}$ tableaux algorithm deciding consistency of $\mathcal{ALC}$-concepts w.r.t. $\mathcal{ALC}$ terminologies [1]. The DL $\mathcal{ALC}$ extends $\mathcal{EL}$ by value restrictions ($\forall$), disjunction ($\sqcup$), and negation ($\neg$). We can decide $C \sqsubseteq^?_{\mathcal{T}} D$ by deciding satisfiability of $C \sqcap \neg D$ w.r.t. $\mathcal{T}$.

The following example presents a general $\mathcal{EL}$-TBox for which the $\mathcal{ALC}$ tableaux algorithm takes exponentially many steps in the worst case. We use the standard $\mathcal{ALC}$ tableaux as describied in [1].

Example 2 For $n \in \mathbb{N}$, let $N_{\mathrm{con}} := \{A, B, C, D\} \cup \{A_i \mid 1 \leq i \leq n\} \cup \{B_i \mid 1 \leq i \leq n\}$ and $N_{\mathrm{role}} := \{r\}$. Define the TBox $\mathcal{T}_n$ as follows:

$$\begin{aligned} C &\doteq A \\ D &\doteq \exists r.B \\ \exists r.B &\sqsubseteq B \\ A &\sqsubseteq \exists r.A \\ \exists r.A_i \sqcap \exists r.B_i &\sqsubseteq B \qquad \text{for every} 1 \leq i \leq n \end{aligned}$$

To be able to apply the tableaux algorithm, the GCIs in $\mathcal{T}_n$ are represented as tautologies:

$$\begin{aligned} &B \sqcup \forall r.\neg B \\ &\neg A \sqcup \exists r.A \\ &B \sqcup \forall r.\neg A_i \sqcup \forall r.\neg B_i \qquad \text{for every} 1 \leq i \leq n \end{aligned}$$

Figure 1 shows (in an abridged way) the first four steps of the tableaux computation for $\mathcal{T}$. The tableaux algorithm starts in Step 0 with a model of one vertex x_0 labeled by $C \sqcap \neg D$. A so-called 'blocking' techniqe is used to avoid the generation of infinitely many vertices for a model: if the label of the new vertex w is a subset of a label of an old vertex v then w is removed, redirecting the edge pointing to w to the old vertex v.

Since x_0 could not be blocked, all GCIs are added to the label of x_0, yielding the situation denoted as Step 1 in Figure 1. In the tableaux, disjunction is dealt with by means of nondeterminism: a GCI of the form $C \sqcup D$ is resolved by nondeterministically choosing between C or D to add to the label set of the vertex under consideration (see [1] for details). Since the concept name A is already contained in the label of x_0, the only possibility to satisfy the GCI $\neg A \sqcup \exists r.A$ (shown boxed in Step 1) is to introduce an r-successor x_1 to x_0. Several other GCIs in the label of x_0 have to be satisfied. In particular, if the algorithm chooses the disjunct $\forall r.\neg B$ from the GCI $B \sqcup \forall r.\neg B$ then $\neg B$ is added to the label set of x_1. Moreover, for every $1 \leq i \leq n$ the GCI

$$B \sqcup \forall r.\neg A_i \sqcup \forall r.\neg B_i \qquad 1 \leq i \leq n$$

must be satisfied for x_0. Since $\neg B$ is already in the label of x_1, thus ruling out choosing B, the algorithm for every i has to include either $\neg A_i$ or $\neg B_i$ into the label of x_1. Hence a set S_1 is added to the label set of x_1, where S_1 corresponds to a tuple $\bar{s}_1$ with

$$\bar{s}_1 \in \{\neg A_1, \neg B_1\} \times \cdots \times \{\neg A_n, \neg B_n\} =: S. \qquad (*)$$

Without going into detail further, Steps 3 and 4 in Figure 1 illustrate that the tableaux algorithm necessarily adds a successor x_2 of x_1 whose label set consists of $A, \neg B$ and another set S_2 representing another nondeterministic choice from S, see $(*)$. Hence, the introduction of x_2 can be blocked only if the algorithm nondeterministically chose $S_1 = S_2$.

Obviously, the situation for x_2 resembles that of x_1, implying that another successor x_3 is introduced and so on. As there exist exponentially many sets S_j mutually incomparable w.r.t. the subset relation the nondeterminism of the tableaux algorithm might give rise to an exponentially long line of successors before a vertex x_k is introduced in whose label the set S_k *necessarily* is a repitition of a label set seen before.

Hence, the standard tableaux algorithm in the worst case needs exponentially many steps to decide the subsumption $C \sqsubseteq_{\mathcal{T}} D$.

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.

Step 0

$$x_0 : \{A \sqcap \forall r.\neg B\} \xrightarrow{\text{GCIs}}$$

Step 1

$$\begin{array}{c} x_0 : \{A \sqcap \forall r.\neg B\} \\ \{\boxed{\neg A \sqcup \exists r.A}, B \sqcup \forall r.\neg B\} \\ \{B \sqcup \forall r.\neg A_i \sqcup \forall r.\neg B_i \mid i\} \end{array} \xrightarrow{\text{add } r\text{-successor}}$$

Step 2

$$\begin{array}{c} x_0 : \{A \sqcap \forall r.\neg B\} \\ \{\neg A \sqcup \exists r.A, B \sqcup \forall r.\neg B\} \\ \{B \sqcup \forall r.\neg A_i \sqcup \forall r.\neg B_i \mid i\} \\ \downarrow r \\ x_1 : \{A, \neg B\} \\ S_1 \end{array}$$

Step 3

$$\begin{array}{c} x_0 : \{A \sqcap \forall r.\neg B\} \\ \{\neg A \sqcup \exists r.A, B \sqcup \forall r.\neg B\} \\ \{B \sqcup \forall r.\neg A_i \sqcup \forall r.\neg B_i \mid i\} \\ \downarrow r \\ x_1 : \{A, \neg B\} \\ \xrightarrow{\text{GCIs}} \{\boxed{\neg A \sqcup \exists r.A}, B \sqcup \forall r.\neg B\} \\ \{\forall r.\neg A_i \sqcup \forall r.\neg B_i \mid i\} \\ S_1 \end{array} \xrightarrow{\text{add } r\text{-successor}}$$

Step 4

$$\begin{array}{c} x_0 : \{A \sqcap \forall r.\neg B\} \\ \{\neg A \sqcup \exists r.A, B \sqcup \forall r.\neg B\} \\ \{B \sqcup \forall r.\neg A_i \sqcup \forall r.\neg B_i \mid i\} \\ \downarrow r \\ x_1 : \{A, \neg B\} \\ \{\neg A \sqcup \exists r.A, B \sqcup \forall r.\neg B\} \\ \{\forall r.\neg A_i \sqcup \forall r.\neg B_i \mid i\} \\ S_1 \\ \downarrow r \\ x_2 : \{A, \neg B\} \\ S_2 \end{array}$$

Figure 1: $\mathcal{ALCI}$ tableaux computation

Definition 3 (Normalized $\mathcal{ELH}$ TBox) Let $\mathcal{T}$ be an $\mathcal{ELH}$-TBox over N_{con} and N_{role}. $\mathcal{T}$ is *normalised* 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 $N_{\text{con}}^{\top}$.

Such a normal form can easily be computed in polynomial time and does not increase the size of the TBox more than polynomially. The following definition provides normalization rules by which an arbitrary $\mathcal{EL}$-TBox can be transformed into a normalized one. The normalization rules are inspired by [18] where a similar problem is solved for $\mathcal{ALC}$-TBoxes containing only definitions.

Definition 4 (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 $N_{\text{con}}^{\mathcal{T},\top}$ and N_{role}, for every $r \in N_{\text{role}}$, and every $\rho \in \{\sqsubseteq, \doteq\}$, the $\mathcal{ELH}$-normalization rules are defined modulo commutativity of conjunction ($\sqcap$) as follows:

$$\begin{array}{lrcl} \textsf{NF1} & \hat{C} \sqcap D \;\rho\; E & \longrightarrow & \{A \doteq \hat{C},\, A \sqcap D \;\rho\; E\} \\ \textsf{NF2} & C \;\rho\; D \sqcap \hat{E} & \longrightarrow & \{C \;\rho\; D \sqcap A,\, A \doteq \hat{E}\} \\ \textsf{NF3} & \exists r.\hat{C} \;\rho\; D & \longrightarrow & \{A \doteq \hat{C},\, \exists r.A \;\rho\; D\} \\ \textsf{NF4} & C \;\rho\; \exists r.\hat{D} & \longrightarrow & \{C \;\rho\; \exists r.A,\, A \doteq \hat{D}\} \\ & & & \\ \textsf{NF5} & C \sqsubseteq D \sqcap E & \longrightarrow & \{C \sqsubseteq D,\, C \sqsubseteq E\} \\ \textsf{NF6} & C \doteq D & \longrightarrow & \{C \sqsubseteq D,\, C \sqsubseteq D\}, \end{array}$$

where $\hat{C}, \hat{D}$ denote concept descriptions that are *no* concept names and A denotes a new concept name from N_{con} not occurring in $\mathcal{T}$. Applying a rule $\mathsf{R} := G \longrightarrow \mathcal{S}$ to $\mathcal{T}$ changes $\mathcal{T}$ to $(\mathcal{T} \setminus \{G\}) \cup \mathcal{S}$. The normalized TBox $norm(\mathcal{T})$ is defined by first exhaustively applying Rules **NF1** to **NF4** and, after that, exhaustively applying Rule **NF5** and, after that, exhaustively applying Rule **NF6**.

The number of possible applications of Rules **NF1** to **NF4** is limited linearly in the size of $\mathcal{T}$. Each of these rules increases the size of $\mathcal{T}$ only by a constant. Hence, applying Rules **NF1** to **NF4** exhaustively increases the size of $\mathcal{T}$ only polynomially. The resulting Tbox may still violate the definition of normalized TBoxes in two respects. Firstly, it may contain GCIs of the form $A \; \rho \; B \sqcap C$ with $A, B, C \in N_{\text{con}}^{\top}$. Secondly, it may contain GCIs with $\rho = \doteq$. As $C \in N_{\text{con}}$, a single application of Rule **NF5** therefore also increases the size of $\mathcal{T}$ only by a constant. Applying the rule exhaustively produces a TBox of linear size in the input. Replacing all GCIs of the $C \doteq D$ by Rule **NF6** obviously has the same property: the size of $\mathcal{T}$ increased only linearly and the rule can be applied only a linear number of times. It is easy to see that normalization also takes only polynomial time.

Our strategy is now, 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 $N_{\text{con}}^{\mathcal{T},\top} := N_{\text{con}}^{\top} \cup \{\top\}$.

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 2. The *implication set* $S_*(A)$ of A is defined as the infinite union $S_*(A) := \bigcup_{i \geq 0} S_i(A)$. Analogously, $S_*(r) := \bigcup_{i \geq 0} S_i(r)$.

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.

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 $A_1 \in S_i(A)$ and $A_1 \sqsubseteq B \in \mathcal{T}$ and $B \notin S_{i+1}(A)$ then $S_{i+1}(A) := S_{i+1}(A) \cup \{B\}$
IS2	If $A_1, A_2 \in S_i(A)$ and $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$ and $B \notin S_{i+1}(A)$ then $S_{i+1}(A) := S_{i+1}(A) \cup \{B\}$
IS3	If $A_1 \in S_i(A)$ 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 $C \notin S_{i+1}(A)$ then $S_{i+1}(A) := S_{i+1}(A) \cup \{C\}$

Figure 2: Rules for implication sets

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

Proof. (i) Proof by induction over n. As $S_0(r) = \{r\}$, the claim holds trivially. For $n > 0$ we know by Rule **ISR** that there exists a role $t \in S_{n-1}(r)$ and a SRI $t \sqsubseteq s \in \mathcal{T}$. By induction hypothesis $r \sqsubseteq_{\mathcal{T}} t$ which by transitivity of role inclusion axioms yields $r \sqsubseteq_{\mathcal{T}} s$. For the reverse direction, $r \sqsubseteq_{\mathcal{T}} s$ immediately implies a finite chain

$$\{r \sqsubseteq t_0\} \cup \{t_i \sqsubseteq t_{i+1} \mid 0 \leq i \leq k-1\} \cup \{t_k \sqsubseteq s\} \subseteq \mathcal{T}$$

of SRIs in $\mathcal{T}$, implying by a finite number of applications of Rule **ISR** that $s \in S_{k+1}(r)$.

(ii) ($\Rightarrow$) It suffices to show for every model $\mathcal{I}$ of $\mathcal{T}$ and for every $B \in S_*(A)$ that $x \in A^{\mathcal{I}}$ implies $x \in B^{\mathcal{I}}$. Assume a model $\mathcal{I}$ of $\mathcal{T}$ with a witness $x \in A^{\mathcal{I}}$ and let $B \in S_*(A)$. Proof by induction over n where n is the least index with $B \in S_n(A)$.

($n = 0$) Then, $S_n(A) = \{A\}$ implying $B = A$. As x was chosen a witness of A the claim holds.

($n > 0$) In Step $n - 1$, B can have been included into $S_n(A)$ by any of the Rules **IS1** to **IS6**. We distinguish one case for each rule.

(**IS1**) There exists a concept name $A_1 \in S_{n-1}(A)$ and a GCI $G := A_1 \sqsubseteq B \in \mathcal{T}$. By induction hypothesis (IH), $x \in A_1^{\mathcal{I}}$, implying by G that also $x \in B^{\mathcal{I}}$.

(**IS2**) There exist two concept names $A_1, A_2 \in S_{n-1}(A)$ and a GCI $G := A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$. By IH, $A_1, A_2 \in S_{n-1}(A)$ yields $x \in A_1^{\mathcal{I}}$ and $x \in A_2^{\mathcal{I}}$, implying by G that $x \in B^{\mathcal{I}}$.

(**IS3**) There exist concept names $A_1 \in S_{n-1}(A)$, $A_2 \in N_{\text{con}}^{\mathcal{T},\top}$, and $A_3 \in S_{n-1}(A_2)$ and two GCIs $G := A_1 \sqsubseteq \exists r.A_2$ and $H := \exists s.A_3 \sqsubseteq B$ with $s \in S_{n-1}(r)$. By IH, $r \sqsubseteq s$, implying by G that $x \in (\exists r.A_2)^{\mathcal{I}}$. Since $A_3 \in S_{n-1}(A_2)$ and , the IH implies $x \in A_1^{\mathcal{I}}$ and $x \in (\exists s.A_3)^I$, yielding by H that $x \in B^{\mathcal{I}}$.

($\Leftarrow$) It suffices to show that if $B \notin S_*(A)$ then we can construct a model $\mathcal{I}$ of $\mathcal{T}$ with a witness $x \in A^{\mathcal{I}} \setminus B^{\mathcal{I}}$.

We construct a (possibly infinite) *canonical model* $\mathcal{I}(A)$ of A w.r.t. $\mathcal{T}$ by means of the following definition. $I(A)$ is defined iteratively starting by $I_0(A)$. Define $\Delta^{\mathcal{I}_0(A)} := \{x_A\}$ and $B^{\mathcal{I}_0(A)} := \{x_A \mid B = A\}$ for all $B \in N_{\text{con}}^{\mathcal{T},\top}$. For $i \geq 0$, the model $\mathcal{I}_{i+1}$ is defined as an extension of $\mathcal{I}_i$ obtained by exhaustive application of the following generation rules.

CM1 If $A \sqsubseteq B \in \mathcal{T}$ then, for every individual $x \in \Delta^{\mathcal{I}_i}$ with $x \in A^{\mathcal{I}_i}$ and $x \notin B^{\mathcal{I}_i}$, add x to $B^{\mathcal{I}_{i+1}}$

CM2 If $A \sqcap B \sqsubseteq C \in \mathcal{T}$ then, for every individual $x \in \Delta^{\mathcal{I}_i}$ with $x \in A^{\mathcal{I}_i} \cap B^{\mathcal{I}_i}$ and $x \notin C^{\mathcal{I}_{i+1}}$, add x to $C^{\mathcal{I}_{i+1}}$

CM3 If $A \sqsubseteq \exists r.B \in \mathcal{T}$ then, for every individual $x \in \Delta^{\mathcal{I}_i}$ with $x \in A^{\mathcal{I}_i}$ for which no r-successor $y \in \Delta^{\mathcal{I}_{i+1}}$ with $y \in B^{\mathcal{I}_{i+1}}$ exists, introduce a new individual y to $\Delta^{\mathcal{I}_{i+1}}$ and include y into $B^{\mathcal{I}_{i+1}}$ and include (x, y) into $r^{\mathcal{I}_{i+1}}$

CM4 If $\exists r.A \sqsubseteq B \in \mathcal{T}$ then, for every pair $(x, y) \in s^{\mathcal{I}_i}$ with $s \sqsubseteq_{\mathcal{T}} r$ and $y \in A^{\mathcal{I}_i}$ and $x \notin B^{\mathcal{I}_{i+1}}$, include x into $B^{\mathcal{I}_{i+1}}$

The above rules are applied fairly, i.e., every rule applicable to already existing elements $x \in \Delta^{\mathcal{I}_i}$ will be applied before applying rules to new elements. The canonical model $\mathcal{I}(A)$ is defined as the infinite union $\mathcal{I}(A) := \bigcup_{i \geq 0} \mathcal{I}_i(A)$.

We first prove that $\mathcal{I}(A)$ in fact is a model of A w.r.t. $\mathcal{T}$. Assume that $x_A \notin A^{\mathcal{I}(A)}$. In this case there is a $y \in \Delta^{\mathcal{I}(A)}$ for which a GCI $G \in \mathcal{T}$ is violated. As $\mathcal{T}$ is normalized, it suffices to distinguish four cases for the violated GCI G.

- If $G = B \sqsubseteq C \in \mathcal{T}$ then $y \in B^{\mathcal{I}(A)}$ but $y \notin C^{\mathcal{I}(A)}$. Consider the least index n with $y \in B^{\mathcal{I}_n(A)}$. By definition, Rule **CM1** causes y to be added to $C^{\mathcal{I}_{n+1}} \subseteq C^{\mathcal{I}}$, contradicting the assumption.

- If $G = B \sqcap C \sqsubseteq D \in \mathcal{T}$ then $y \in B^{\mathcal{I}(A)} \sqcap C^{\mathcal{I}(A)}$ but $y \notin D^{\mathcal{I}(A)}$. Consider the least index n with $y \in B^{\mathcal{I}_n(A)} \sqcap C^{\mathcal{I}_n(A)}$. Rule **CM2** causes y to be added to $D^{\mathcal{I}_{n+1}} \subseteq D^{\mathcal{I}}$, in contradiction to the assumption.

- If $G = B \sqsubseteq \exists r.C \in \mathcal{T}$ then $y \in B^{\mathcal{I}(A)}$ but y has no appropriate r-successor. Consider the least n with $y \in B^{\mathcal{I}_n(A)}$. By Rule **CM3**, a new element z is introduced to $\Delta^{\mathcal{I}_{n+1}}$, the pair (y, z) added to $r^{\mathcal{I}_{n+1}}$, and z added to $C^{\mathcal{I}_{n+1}}$, again in contradiction to the assumption.

- If $G = \exists r.B \sqsubseteq C \in \mathcal{T}$ then there exists an edge $(y, z) \in s^{\mathcal{I}(A)}$ with $s \sqsubseteq_{\mathcal{T}} r$ such that $z \in B^{\mathcal{I}(A)}$ but $y \notin C^{\mathcal{I}(A)}$. Consider the least n with $z \in B^{\mathcal{I}_n(A)}$. As $s \sqsubseteq_{\mathcal{T}} r$ and $(y, z) \in s^{\mathcal{I}_n(A)}$, Rule **CM4** adds y to $C^{\mathcal{I}_{n+1}(A)} \subseteq C^{\mathcal{I}(A)}$, contradicting the assumption.

Having proven $\mathcal{I}(A)$ to be a model of A w.r.t. $\mathcal{T}$ it remains to show that $B^{\mathcal{I}(A)} \not\subseteq A^{\mathcal{I}(A)}$. To this end, we show for every $n \in \mathbb{N}$, for *every* $A, B \in N_{\text{con}}^{\mathcal{T},\top}$, $A \neq B$, and for *every* $x \in A^{\mathcal{I}_n(A)}$: if $\{C \mid C \in x^{\mathcal{I}_t(A)}\} = \{A\}$ for some minimally chosen $t \in \mathbb{N}$ and $x \in B^{\mathcal{I}_n(A)}$ then $B \in S_*(A)$. Note that $B \in S_*(A)$ holds if $B \in S_m(A)$ for some $m \in \mathbb{N}$ since $S_m(A) \subseteq S_*(A)$.

$(n = 0)$ Trivial since $B^{\mathcal{I}_0(A)} = \emptyset$ implies that the premise $x \in B^{\mathcal{I}_n(A)}$ does not hold.

$(n \geq 0)$ Let $\{C \mid C \in x^{\mathcal{I}_t(A)}\} = \{A\}$ for some $t < n$ and let $x \in B^{\mathcal{I}_n(A)} \setminus B^{\mathcal{I}_{n-1}(A)}$. In the definition of $\mathcal{I}_n(A)$ there are four rules which can have caused the inclusion of x into $B^{\mathcal{I}_n(A)}$:

- (**CM1**) Then then there is a GCI $G := A_1 \sqsubseteq B \in \mathcal{T}$ and $x \in A_1^{\mathcal{I}_{n-1}(A)}$. If $t = n - 1$ then $A_1 = A$, implying $B \in S_1(A)$ by Rule **IS1** with G. If $t < n - 1$ then, by induction hypothesis (IH), $A_1 \in S_*(A)$, implying $A_1 \in S_m(A)$ for some $m \in \mathbb{N}$, yielding $B \in S_{m+1}(A)$ by Rule **IS1** with G.

- (**CM2**) Then there is a GCI $G := A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$ and $x \in A_1^{\mathcal{I}_{n-1}(A)} \cap A_2^{\mathcal{I}_{n-1}(A)}$. If $t = n - 1$ then $A_1 = A_2 = A$, implying $B \in S_1(A)$ by Rule **IS2** with G. If $t < n - 1$ then, by IH, $\{A_1, A_2\} \subseteq S_*(A)$. Hence, $\{A_1, A_2\} \subseteq S_m(A)$ for some $m \in \mathbb{N}$, implying $B \in S_{m+1}(A)$ by Rule **IS2** with G.

- (**CM4**) Then there is a GCI $G := \exists r.A_1 \sqsubseteq B \in \mathcal{T}$ and $y \in \Delta^{\mathcal{I}_{n-1}(A)}$ with $(x, y) \in s^{\mathcal{I}_{n-1}(A)}$ with $s \sqsubseteq_{\mathcal{T}} r$ and $y \in A_1^{\mathcal{I}_{n-1}(A)}$, implying $t < n-1$ since x and y cannot be created at the same time. Hence, firstly, there is a

GCI $H := C \sqsubseteq \exists s.D \in \mathcal{T}$ and an index $t \leq k < n-1$ with $x \in C^{\mathcal{I}_k}$, implying $(x,y) \in s^{\mathcal{I}_{k+1}(A)}$ and $y \in D^{\mathcal{I}_{k+1}}$. Secondly, $y \in A_1^{\mathcal{I}_{k+1}}$. By IH, $A_1 \in S_*(D)$. If $t = k$ then $C = A$, otherwise, by IH, $C \in S_*(A)$. In both cases there exists a least index m with $C \in S_m(A)$ and $A_1 \in S_m(D)$, implying $B \in S_{m+1}(A)$ by Rule **IS3** with G and H.

∎

We have shown how to decide subsumption w.r.t. general $\mathcal{ELH}$-TBoxes. It remains to show that our decision procedure works in polynomial time. In contrast to the correctness proof this is relatively easy.

Lemma 7 *For every normalised $\mathcal{ELH}$-TBox over N_{con} and N_{role} and for every $A \in N_{\text{con}}^{\mathcal{T},\top}$, the implication set $\mathcal{S}_*(A)$ can be computed in polynomial time in the size of $\mathcal{T}$.*

PROOF. 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, 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 8 *Subsumption in $\mathcal{ELH}$ w.r.t. GCIs can be decided in polynomial time.*

4 Co-NP hard extensions

The surprisingly low upper bound for the subsumption problem in $\mathcal{ELH}$ w.r.t. general TBoxes gives rise to the question whether it might be possible to extend $\mathcal{ELH}$ by other constructors without losing polynomiality. From a knowledge representaton perspective, particularly useful constructors might be number restrictions $(\leq n\, r)$ and $(\geq n\, r)$, and disjunction $(\sqcup)$. The DL K-REP [9] provides the constructor 'allsome' $(\forall\exists)$ to capture the meaning often associated with 'for all' statements in natural language. A concept $\forall\exists.C$ is equivalent to $\forall.C \sqcap \exists r.C$. A a value restriction $\forall.C$ alone cannot be expressed by means of allsome.

In the following sections we show that adding one of the constructors number restriction, disjunction, and allsome makes the subsumption problem co-NP hard—even without GCIs. In case of number restriction and disjunction (Sections 4.1 and 4.2, resp.), co-NP hardness holds even for subsumption w.r.t. the empty TBox. In case of allsome (Section 4.3), the lower bound holds already for acyclic TBoxes without GCIs or SRIs.

4.1 $\mathcal{EL}$ + number restriction

We show co-NP hardness of the subsumption problem in $\mathcal{ELN}$ by reducing BIN-PACKING to consistency of $\mathcal{ELN}$ concepts. Since $\mathcal{ELN}$ can express inconsistency as $(\leq 0\, r) \sqcap (\geq 1\, r)$, inconsistency can be reduced to non-subsumption of $\mathcal{ELN}$ concepts, yielding the desired reduction.

Definition 9 (BIN-PACKING) Let U be a nonempty finite set. Let $s\colon U \to \mathbb{N}^+$ and let $b, k \in \mathbb{N}^+$. Then, $P := (U, s, b, k)$ is a *Bin-Packing* problem. A *solution* to P is a partition of U into k pairwise disjoint sets $U_1, \ldots, U_k$ such that for all $i \in \{1, \ldots, k\}$ it holds that $\Sigma_{u \in U_i} s(u) \leq b$.

BIN-PACKING is an NP-complete problem in the strong sense [11, p. 226], implying that we may assume unary encoding for the numbers in P. Given P, we construct a concept C_P which is satisfiable iff P has a solution.

The intuition behind C_P is to use a concept description of fixed depth 2 and, (i) express on toplevel that at most k bins, i.e., k pairwise disjoint sets $U_1, \ldots, U_k$, exist, (ii) express on the first role level that every bin weighs at most b, and (iii) use the second role level to represent the weights $s(u)$ of the objects $u \in U$. The following definition formalizes this notion.

Definition 10 (Bin-packing concept) Let $P = (U, s, b, k)$ be a Bin-Packing problem. Let $\ell := \lceil \lg(\Sigma_{u \in U} s(u)) \rceil$. Define $N_{\text{prim}}^P := \emptyset$ and $N_{\text{role}}^P := \{r\} \cup \{r_1, \ldots, r_\ell\}$. Let

$$\mathcal{C}^P := \left\{ \bigsqcap_{i=1}^{\ell} C_i \;\middle|\; C_i \in \{(\leq 0\, r_i), (\geq 1\, r_i)\} \right\}$$

Let $f^P\colon \{(u,i) \mid u \in U, 1 \leq i \leq s(u)\} \to \mathcal{C}^P$ be an injective mapping. The $\mathcal{ELN}$-concept description C^P is defined as follows:

$$C^P := (\leq k\, r) \sqcap \bigsqcap_{u\in U} \exists r. \left((\leq b\, r) \sqcap \bigsqcap_{i=1}^{s(u)} \exists r. f^P(u,i) \right)$$

Note that $\Sigma_{u\in U} s(u) \leq |\mathcal{C}^P| < 2\cdot\Sigma_{u\in U} s(u)$ so that f^P in fact exists and can be computed easily in polynomial time in the size of P with unary number encoding. The above definition is well-defined only w.r.t. the mapping f^P of which in general many different ones exist. Nevertheless, for our purpose an arbitrary but fixed instance of f^P suffices.

The motivation behind the function $f(u,i)$ is to provide a simple method to count binarily from 0 to $\sigma_{u\in U} s(u)$, the sum of the weights of all elements in U. The concept description $f(u,i)$ on the second role level, i.e., in the leaves of C_P, enforce that no two leaves can be represented by the same element in a model of C_P. This is guaranteed that by the fact that two arbitrary but different leaves in C_P differ in negating or not negating at least one number restriction for a role r_i. The following lemma proves formally that the reduction is correct.

Lemma 11 *Let $P = (U, s, b, k)$ be a Bin-Packing problem and C^P the corresponding concept description over N^P_{prim} and N^P_{role}. Then,*

1. *For every $u, v \in U$, $i \in \{1, \dots, s(u)\}$, and $j \in \{1, \dots, s(v)\}$ it holds that $f^P(u,i) \sqcap f^P(v,j) \equiv \bot$ iff $u \neq v$ or $i \neq j$.*

2. *P has a solution iff C^P is satisfiable.*

PROOF. $(1, \Leftarrow)$ If $u = v$ and $i = j$ then $f^P(u,i) \sqcap f^P(v,j) \equiv f^P(u,i) \in \mathcal{C}^P$. Every concept description in $\mathcal{C}^P$ is consistent because each of its conjuncts imposes a number restriction on a different role.

$(1, \Rightarrow)$ Then the injectivity of f^P implies that $f^P(u,i)$ and $f^P(v,j)$ are two distinct concepts in $\mathcal{C}^P$. Hence, there exists an index t such that $f^P(u,i)$ contains the conjunct $(\leq 0\; r_t)$ and $f^P(u,i)$ contains the conjunct $(\geq 1\; r_t)$ or vice versa. Hence, the conjunction $f^P(u,i) \sqcap f^P(v,j)$ is subsumed by $(\leq 0\, r_t) \sqcap (\geq 1\, r_t) \equiv \bot$.

$(2, \Rightarrow)$ Denote by $U_1, \dots, U_k$ a solution to P. Define a model $\mathcal{I}$ of C^P as follows:

$$\Delta^{\mathcal{I}} := \{w, z\} \cup \{x_i \mid 1 \leq i \leq k\} \cup \{y_{uj} \mid u \in U,\ 1 \leq j \leq s(u)\}.$$

Let

$$r^{\mathcal{I}} := \bigcup_{i=1}^{k} \big(\{(w, x_i)\} \cup \{(x_i, y_{uj}) \mid u \in U_i,\ 1 \le j \le s(u)\}\big)$$

and for every $t \in \{1, \ldots, \ell\}$, let

$$r_t^{\mathcal{I}} := \{(y_{uj}, z) \mid u \in U,\ 1 \le j \le s(u),\ f^P(u, j) \sqsubseteq (\ge 1\, r_t)\}.$$

We show that $w \in C^{P^{\mathcal{I}}}$, i.e., w is a witness of C^P. The definition of $r^{\mathcal{I}}$ shows that w has exactly k successors w.r.t. the role r, namely $x_1, \ldots, x_k$. Hence, the number restriction on the toplevel of C^P is satisfied. For the rest of C^P, consider an arbitrary $u \in U$ and select $i \in \{1, \ldots, k\}$ such that $u \in U_i$. It suffices to show that $x_i \in (\le b\ r)^{\mathcal{I}}$ and that $x_i \in (\exists r.f^P(u, j))^{\mathcal{I}}$ for all $1 \le j \le s(u)$.

Due to the definition of $r^{\mathcal{I}}$, x_i has exactly one successor y_{ui} for every element $u \in U_i$ and for every $1 \le j \le s(u)$. Hence, the total number of successors of x_i equals $\Sigma_{u \in U_i} s(u)$ which does not exceed b, the size limit for every U_i.

Consider an arbitrary $j \in \{1, \ldots, s(u)$. Since $(x_i, y_{uj}) \in R^{\mathcal{I}}$ it suffices to show that $y_{uj} \in f^P(u, j)^{\mathcal{I}}$. By definition, $f(u, j) = \sqcap_{t=1}^{\ell} C_t$ with $C_t = (\le 0 r_t)$ or $C_t = (\ge 1 r_t)$ for every $t \in \{1, \ldots, \ell\}$. For every t, the pair (y_{uj}, z) occurs in $r_t^{\mathcal{I}}$ iff $C_t = (\ge 1\, r_t)$. Hence y_{uj} has no successor w.r.t. every role r_t occurring in a number restriction $(\le 0\ r_t)$ and has t as successor w.r.t. every role r_t occurring in a number restriction $(\ge 1\, r_t)$. Thus, y_{uj} is a witness of $f^P(u, j)$.

$(2, \Leftarrow)$ To ease notation, for all $u \in U$ let

$$C^u := (\le b\ r) \sqcap \bigsqcap_{j=1}^{s(u)} \exists r.f^P(u, j).$$

Hence, C^P can be written as

$$C^P = (\le k\ r) \sqcap \bigsqcap_{u \in U} \exists r.C^u.$$

Denote by $\mathcal{I}$ a model of C^P and denote by w a witness $w \in C^{P^{\mathcal{I}}}$. Due to the number restriction on the toplevel of C^P, w has at most k successors w.r.t. r. The $|U|$ existential restrictions on the other hand guarantee that at least one successor exists. Denote by $X := \{x_1, \ldots, x_{k'}\}$ the set of r-successors of w. If $k' < k$ then w.l.o.g., $k - k'$ isolated vertices $x_{k'+1}, \ldots, x_k$ may be added to $\Delta^{\mathcal{I}}$.

Define the partition of U as follows: starting from 1, for $i = 1, \ldots, k$ let

$$U_i := \{u \in U \mid x_i \in C^{u\mathcal{I}},\ \forall j < i\colon u \notin U_j\}.$$

Note that the above definition is well-defined only w.r.t. an order on $\{1, \ldots, k\}$ by which to compute the U_i. We have to show that $U_1, \ldots, U_k$ in fact is a partition of U and that for every $1 \leq i \leq k$ the overall size $\Sigma_{u \in U_i} s(u)$ does not exceed b.

As $w \in C^{P\mathcal{I}}$, every C^u must have a witness in the set X. Thus, the union over all subsets U_i yields U. The restriction $u \notin U_j$ in the definition of every U_i ensures that for every $u \in U$ at most one index i exists with $u \in U_i$. Hence, $U_1, \ldots, U_k$ is a partition of U.

Let $i \in \{1, \ldots, k\}$. By definition, U_i contains a subset of all $u \in U$ of which x_i is a witness. If U_i is nonempty, then two facts are implied. Firstly, x_i has at most b successors w.r.t. r because of the number restriction in one C_u. Secondly, x_i has at least $\sigma_{u \in U_i} s(u)$ successors w.r.t. r. This holds due to the existential restrictions of the form $\exists r.f^P(u, j)$ in every C^u with $u \in U_i$: for every $u \in U_i$ and for every $j \in \{1, \ldots, s(u)\}$, denote by y_{uj} the r-successor of x_i implied by $\exists r.f^P(u, j)$. Assume that $y_{uj} = y_{u'j'}$ for some $u, v \in U_i$, $j \in \{1, \ldots, s(u)\}$, and $j' \in \{1, \ldots, s(u)\}$. Then, y_{uj} is a witness of $f^P(u, j) \sqcap f^P(u', j')$, in contradiction to Claim (1) of the proof. ∎

As satisfiability of $\mathcal{ELN}$ concepts can be reduced to subsumption, i.e., a concept description C is satisfiable if and only if $C \not\sqsubseteq \bot \equiv (\leq 0\, r) \sqcap (\geq 1\, r)$, we immediately obtain the following hardness results:

Corollary 12 *Deciding satisfiability in $\mathcal{ELN}$ w.r.t. the empty TBox is NP-hard. Deciding subsumption in $\mathcal{ELN}$ w.r.t. the empty TBox is co-NP-hard.*

4.2 $\mathcal{EL}$ + disjunction

We show co-NP hardness of the subsumption problem in $\mathcal{ELU}$ by reducing MONOTONE 3SAT to non-subsumption of $\mathcal{ELU}$-concept descriptions. The monotone problem differs from 3SAT only in that every clause contains either only negated or only unnegated literals.

Definition 13 (MONOTONE 3SAT) Let U be a set of variables and S^+, S^- be two sets of clauses over U such that every $s \in S^+$ contains exactly 3 un-negated variables and every $s \in S^-$ exactly 3 negated ones. Then, $P :=$

(U, S^+, S^-) is called a *Monotone 3Sat* problem. A *solution* to P is a truth assignment $t\colon U \to \{0,1\}$ satisfying $S^+ \cup S^-$.

MONOTONE 3SAT is an NP-complete problem [11, p. 259]. We can immediately represent the clauses in S^+ and S^- in $\mathcal{ELU}_\neg$, an extension of $\mathcal{ELU}$ by atomic negation. The conjunction over all clauses is then split into $C \sqcap D$, C containing all positive clauses and D all negative ones. Satisfiability of $C \sqcap D$ is reduced to $\mathcal{ELU}$-non-subsumption by deciding $C \not\sqsubseteq \mathrm{nnf}(\neg D)$ where nnf denotes the negation normal form of $\neg D$. The following lemma provides the formal proof.

Lemma 14 *Let $P = (U, S^+, S^-)$ be a Monotone 3Sat problem. Then there exist $\mathcal{ELU}$-concept descriptions C, D such that P has a solution iff $C \not\sqsubseteq D$.*

PROOF. Let $N_{\mathrm{prim}} := U$ and $N_{\mathrm{role}} := \emptyset$. We can immediately translate $S^+ \cup S^-$ into an $\mathcal{ELU}_\neg$-concept description C^P of the following form:

$$C^P := \bigsqcap_{s \in S^+} \bigsqcup_{u \in s} u \sqcap \bigsqcap_{s \in S^-} \bigsqcup_{\neg u \in s} \neg u.$$

It is easy to see that P has a solution iff C^P is satisfiable. The satisfiability of C^P is eqiuvalent to the non-subsumption

$$C := \bigsqcap_{s \in S^+} \bigsqcup_{u \in s} u \not\sqsubseteq \neg \bigsqcap_{s \in S^-} \bigsqcup_{\neg u \in s} \neg u \equiv \bigsqcup_{s \in S^-} \bigsqcap_{\neg u \in s} u =: D.$$

Observe that both C and D are concept descriptions in $\mathcal{ELU}$. ∎

Corollary 15 *Deciding subsumption of $\mathcal{ELU}$-concept descriptions w.r.t. the empty TBox is co-NP-hard.*

The above reduction implies co-NP-hardness of the subsumption problem even for the very small description logic providing only conjunction and disjunction.

4.3 $\mathcal{EL}$ + allsome

We show co-NP hardness of subsumption in $\mathcal{EL}_{\forall\exists}$ by reduction of the subsumption problem in $\mathcal{FL}_0$ w.r.t. acyclic simple terminologies to the analogous problem in $\mathcal{L}_{\forall\exists}$, a sublanguage of $\mathcal{EL}_{\forall\exists}$ without existential restrictions. The first problem is known to be co-NP hard.

Our aim is to translate acyclic simple $\mathcal{FL}_0$-TBoxes, i.e., containing no GCIs or SRIs, into subsumption-preserving equivalent ones over $\mathcal{L}_{\forall\exists}$, thereby reducing the subsumption problem from one DL to the other. To this end, we introduce a normal form for $\mathcal{FL}_0$-TBoxes that simplifies the translation.

Definition 16 (translation function) Let $\mathcal{T}$ be an arbitrary $\mathcal{FL}_0$-TBox over N_{con}, and N_{role}. $\mathcal{T}$ is called *reduced* iff none of the following transformation rules can be applied to any concept description D with $C \doteq D \in \mathcal{T}$ or any of its subdescriptions:

$$\begin{aligned} \forall r.\top &\longrightarrow \top \\ E &\longrightarrow \top \quad \text{iff } E \doteq \top \in \mathcal{T} \\ F \sqcap \top &\longrightarrow F, \end{aligned}$$

where $r \in N^{\mathcal{T}}_{\text{role}}$, E represents an arbitrary defined concept, and F an arbitrary concept description over $N^{\mathcal{T}}_{\text{con}}$, and $N^{\mathcal{T}}_{\text{role}}$. For a reduced TBox $\mathcal{T}$, the translated TBox $\mathit{trans}(\mathcal{T})$ is defined by syntactically replacing all $\forall$-quantors by $\forall\exists$-quantors: $\mathit{trans}(\mathcal{T}) := \mathcal{T}\{\forall/\forall\exists\}$.

Note that the above definition is correct only in the sense that all subsumption relations are preserved. While a model of $\mathit{trans}(\mathcal{T})$ can always be shown to be model of $\mathcal{T}$, the reverse direction need *not* hold.

To prove correctness of the translation we first devise a formal-language characterization of subsumption for $\mathcal{L}_{\forall\exists}$-concept descriptions. Note that we may restrict our attention to subsumption w.r.t. the empty TBox since acyclic TBoxes can be expanded until no defined concepts occur on right-hand sides of concept definitions. In $\mathcal{FL}_0$, the equivalence $\forall r.(C \sqcap D) \equiv \forall r.C \sqcap \forall r.D$ gives rise to a particularly simple representation of concept descriptions, called *unfolding* in [21] or *concept centered normal form* in [3]. Given a concept description C, the idea is to exploit the above equivalence from left to right until conjunction in C occurs only on toplevel, implying that all value restrictions are of the form $\forall r_1.\forall r_2.\cdots\forall r_n.A$ with $A \in N_{\text{prim}}$. The word $r_1 r_2 \ldots r_n$ can then be used to represent the corresponding restriction C imposes w.r.t. A.

The same principle holds for $\mathcal{L}_{\forall\exists}$: a concept description $\forall\exists r.(C \sqcap D)$ by definition equals $\forall r.(C \sqcap D) \sqcap \exists r.(C \sqcap D)$. Because of the propagation from value to existential restrictions, replacing $\exists r.(C \sqcap D)$ by $\exists r.\top$ preserves equivalence. Duplicating $\exists r.\top$, the propagation argument in the reverse direction yields $\forall\exists r.C \sqcap \forall\exists r.D$. We will use the above normalization to define so-called *role languages* for atomic concepts occurring in concept descriptions. To ease notation, we start by extending the constructors $\forall$ and $\forall\exists$ to words over N_{role}. In the remainder of this section, we may w.l.o.g. assume that N_{con} and N_{role} are not only finite but limited by the concept names and role names occurring in the concept descriptions C, D for which we want to decide $C \sqsubseteq D$.

Definition 17 (Word restrictions) For all $A \in N_{\text{prim}}$, $r \in N_{\text{role}}$, $w \in N_{\text{role}}^*$, and for $Q \in \{\forall, \forall\exists\}$, the concept description $Qw.A$ is inductively defined by:

$$\begin{aligned} Q\varepsilon.A &:= A \\ Qrw.A &:= Qr.Qw.A \end{aligned}$$

As we need to refer to the (already existing) role-language characterization for $\mathcal{FL}_0$, we simultaneously introduce role languages for $\mathcal{FL}_0$-concept descriptions and $\mathcal{L}_{\forall\exists}$-concept descriptions. Obviously, while $\top$ can be ignored for $\mathcal{FL}_0$ it must be treated as an ordinary concept name in $\mathcal{L}_{\forall\exists}$.

Definition 18 (Role languages) Let C be an $\mathcal{FL}_0$-concept description. Then, for $Q = \forall$ and for arbitrary $A, B \in N_{\text{prim}}$ the formal language $L_A(C)$ is inductively defined by:

$$\begin{aligned} L_A(\top) &:= \emptyset \\ L_A(B) &:= \{\varepsilon \mid A = B\} \\ L_A(\sqcap_i C_i) &:= \bigcup_i L_A(C_i) \\ L_A(Qr.C) &:= \{r\} \cdot L_A(C) \end{aligned}$$

For $\mathcal{L}_{\forall\exists}$-concept descriptions ($Q = \forall\exists$) the top-concept $\top$ is treated like a primitive concept. Hence, the inductive definition is extended to arbitrary $A, B \in N_{\text{prim}} \cup \{\top\}$ and the definition $L_A(\top) := \emptyset$ is removed.

The language $L_A(C)$ contains all words $r_1 \ldots r_n$ over N_{role} with $C \sqsubseteq Qr_1. \cdots Qr_n.A$, where $Q = \forall$ in case of $\mathcal{FL}_0$ and $Q = \forall\exists$ in case of $\mathcal{L}_{\forall\exists}$. In [21] it was shown that the set of all role languages of a given $\mathcal{FL}_0$-concept description in fact *characterizes* the concept up to equivalence. The following lemma holds:

Lemma 19 *Let C be an $\mathcal{FL}_0$-concept description over N_{prim} and N_{role}. Then,* $C \equiv \bigsqcap_{A \in N_{\mathrm{prim}}} \bigsqcap_{w \in L_A(C)} \forall w.A$

In [2], subsumption of $\mathcal{FL}_\perp$ concept descriptions $C \sqsubseteq D$ was characterized by the role languages of C and D. For the sublanguage $\mathcal{FL}_0$, we can immediately derive the following characterization of subsumption of $\mathcal{FL}_0$-concept descriptions.

Lemma 20 *Let C, D be $\mathcal{FL}_0$-concept descriptions over N_{prim} and N_{role}. Then, $C \sqsubseteq D$ iff $L_A(C) \supseteq L_A(D)$ for all $A \in N_{\mathrm{role}}$.*

We aim at a similar characterization of subsumption for $\mathcal{L}_{\forall\exists}$. Therefore, our first step is to prove that $\mathcal{L}_{\forall\exists}$-concept descriptions can in fact also be characterized by their role languages.

Lemma 21 *Let C, D be $\mathcal{L}_{\forall\exists}$-concept descriptions over N_{prim} and N_{role} and let $r \in N_{\mathrm{role}}$. Then, $\forall\exists r.(C \sqcap D) \equiv \forall\exists r.C \sqcap \forall\exists r.D$*

Proof. Due to the semantics of $\forall\exists$-restrictions it is easy to transform $\mathcal{L}_{\forall\exists}$-concept descriptions into equivalent $\mathcal{FLE}$-concept descriptions: it holds that $\forall\exists r.C$ is equivalent to $\forall r.C \sqcap \exists r.\top$. Hence, $\forall\exists r.(C \sqcap D)$ is equivalent to $\forall r.(C \sqcap D) \sqcap \exists r.\top$ which again is equivalent to $\forall r.C \sqcap \forall r.D \sqcap \exists r.\top \sqcap \exists r.\top$ which can be simplified to $\forall\exists r.C \sqcap \forall\exists r.D$. ∎

This immediately yields the extension of Lemma 19 to $\mathcal{L}_{\forall\exists}$-concept descriptions. Note that in contrast to the analogous case of $\mathcal{FL}_0$, in $\mathcal{L}_{\forall\exists}$ allsome restrictions containing only $\top$ are not irrelevant.

Lemma 22 *Let C be an $\mathcal{L}_{\forall\exists}$-concept description over N_{prim} and N_{role}. Then,* $C \equiv \bigsqcap_{A \in N_{\mathrm{prim}}} \bigsqcap_{w \in L_A(C)} \forall\exists w.A \sqcap \bigsqcap_{w \in L_\top(C)} \forall\exists w.\top.$

The following lemma provides a role-language characterization of subsumption of $\mathcal{L}_{\forall\exists}$ concept descriptions w.r.t. the empty TBox. Obviously, the interesting part is the treatment of the $\top$-concept for which an additional equation is introduced.

Lemma 23 *Let C, D be $\mathcal{L}_{\forall\exists}$-concept descriptions over N_{prim} and N_{role}. Then, $C \sqsubseteq D$ iff*

1. *$L_A(C) \supseteq L_A(D)$ for all $A \in N_{\text{role}}$; and*

2. *$L_\top(C) \cup \bigcup_{A \in N_{\text{prim}}} L_A(C) \cup \{\varepsilon\} \supseteq L_\top(D)$.*

PROOF. ($\Rightarrow$) Assume that one of the above conditions is violated. In the first case there exists an atomic concept $A \in N_{\text{prim}}$ with $L_A(C) \not\supseteq L_A(D)$. Hence, there is a word $w \in L_A(D)$ such that $w \notin L_A(C)$. We can now construct a model $\mathcal{I}$ of C which is no model of D. Let $\Delta^\mathcal{I} := \{a_0, \ldots, a_{|w|+1}\}$. Let $A^\mathcal{I} := \Delta^\mathcal{I} \setminus \{a_{|w|}\}$. For all $B \in N_{\text{prim}} \setminus \{A\}$, let $B^\mathcal{I} := \Delta^\mathcal{I}$. For all $r \in N_{\text{role}}$, define

$$r^\mathcal{I} := \{(a_i, a_{i+1}) \mid 0 \leq i \leq |w|\} \cup \{(a_{|w|+1}, a_{|w|+1})\}.$$

The model $\mathcal{I}$ is constructed in such a way that every vertex a_i has a successor w.r.t. every role $r \in N_{\text{role}}$. Moreover, every vertex except $a_{|w|}$ is a witness of all atomic concepts $B \in N_{\text{prim}}$. Only $a_{|w|}$ is a witness of all atomic concepts except A.

Since $w \notin L_A(C)$ ist is easy to see that a_0 is a witness of C but not a witness of D, where a w-chain of successors must lead to a witness of A. This contradicts $C^\mathcal{I} \subseteq D^\mathcal{I}$ as implied by $C \sqsubseteq D$.

If the second condition is violated, we similarly find a word $w \in L_\top(D) \setminus \{\varepsilon\}$ with $w \notin L_\top(C)$ and $w \notin L_A(C)$ for every $A \in N_{\text{prim}}$. Since $|w| \geq 1$ we may write w as vs with $s \in N_{\text{role}}$. Let $\Delta^\mathcal{J} := \{a_0, \ldots, a_{|w|}\}$. For all $A \in N_{\text{prim}}$, let $A^\mathcal{J} := \Delta^\mathcal{J}$, i.e., all atomic concepts hold in every vertex of $\mathcal{J}$. For all $r \in N_{\text{role}}$, define

$$r^\mathcal{J} := \{(a_i, a_{i+1}) \mid 0 \leq i \leq |v| - 1\} \cup \{(a_{|v|}, a_{|v|+1}) \mid r \neq s\} \cup \{(a_{|v|+1}, a_{|v|+1})\}.$$

In the model $\mathcal{J}$ every vertex a_i except $a_{|v|}$ has a successor w.r.t. every role $r \in N_{\text{role}}$. The vertex $a_{|v|}$ has a successor w.r.t. every role except s. It is therefore easy to see that a_0 is a witness of C but none of D where an s-successor must be present after travelling a v-path. This contradicts $C^\mathcal{I} \subseteq D^\mathcal{I}$.

($\Leftarrow$) Then we know that C is equivalent to

$$\bigsqcap_{A \in N_{\text{prim}}} \bigsqcap_{w \in L_A(C)} \forall\exists w.A \;\sqcap \bigsqcap_{w \in L_\top(C)} \forall\exists w.\top$$

and analogously for D. Using the subset relations from Condition 1, we can write C as

$$\bigsqcap_{A \in N_{\text{prim}}} \bigsqcap_{w \in L_A(D)} \forall\exists w.A \;\sqcap \bigsqcap_{A \in N_{\text{prim}}} \bigsqcap_{w \in L_A(C)} \forall\exists w.A \;\sqcap \bigsqcap_{w \in L_\top(C)} \forall\exists w.\top.$$

Since $\forall\exists w.A \sqsubset \forall\exists w.\top$ we may (1) add subdescriptions $\forall\exists w.\top$ for which w also occurs in a subdescription referring to some $A \in N_{\text{prim}}$; and (2) add $\top$.

$$\begin{aligned}&\prod_{A\in N_{\text{prim}}}\prod_{w\in L_A(D)}\forall\exists w.A \sqcap \prod_{A\in N_{\text{prim}}}\prod_{w\in L_A(C)}\forall\exists w.A\\ &\sqcap \prod_{w\in L_\top(C)}\forall\exists w.\top \sqcap \prod_{w\in\bigcup_{A\in N_{\text{prim}}\cup\{\varepsilon\}}L_A(C)}\forall\exists w.\top\end{aligned}$$

Exploiting the subset relation in Condition 2 the concept C can be further rewritten to

$$\begin{aligned}&\prod_{A\in N_{\text{prim}}}\prod_{w\in L_A(D)}\forall\exists w.A \sqcap \prod_{A\in N_{\text{prim}}}\prod_{w\in L_A(C)}\forall\exists w.A\\ &\sqcap \prod_{w\in L_\top(D)}\forall\exists w.\top \sqcap \prod_{w\in L_\top(C)}\forall\exists w.\top \sqcap \prod_{w\in\bigcup_{A\in N_{\text{prim}}\cup\{\varepsilon\}}L_A(C)}\forall\exists w.\top\end{aligned}$$

which equals

$$\begin{aligned}D \sqcap &\prod_{A\in N_{\text{prim}}}\prod_{w\in L_A(C)}\forall\exists w.A\\ &\sqcap \prod_{w\in L_\top(C)}\forall\exists w.\top \sqcap \prod_{w\in\bigcup_{A\in N_{\text{prim}}\cup\{\varepsilon\}}L_A(C)}\forall\exists w.\top.\end{aligned}$$

Hence, C is equivalent to or more specific than D, i.e., $C \sqsubseteq D$. ∎

The above characterizaton of subsumption allows a straightforward proof of correctness of the translation from $\mathcal{FL}_0$ to $\mathcal{L}_{\forall\exists}$.

Lemma 24 *Let $\mathcal{T}$ be an acyclic reduced $\mathcal{FL}_0$-TBox over N_{con}, and N_{role}. Let $A, B \in N_{\text{def}}$. Then, $A \sqsubseteq_{\mathcal{T}} B$ iff $A \sqsubseteq_{\mathit{trans}(\mathcal{T})} B$.*

PROOF. Let $A \doteq C$ and $B \doteq D \in \mathcal{T}$. Denote by $\tilde{\mathcal{T}}$ the TBox resulting from unfolding[2] the original TBox $\mathcal{T}$. In $\tilde{\mathcal{T}}$ no defined concepts occur on any right-hand side of a concept definition. Denote by $\tilde{C}_0, \tilde{D}_0$ the concept descriptions C and D, respectively, from the unfolded TBox $\tilde{C}$. Analogously, denote by $\tilde{C}_{\forall\exists}, \tilde{D}_{\forall\exists}$ the concept descriptions C and D, respectively, from the translated and unfolded TBox $\widetilde{\mathit{trans}}(C)$. It is easy to see that $A \sqsubseteq_{\mathcal{T}} B$ iff $\tilde{C}_0 \sqsubseteq \tilde{D}_0$ and $A \sqsubseteq_{\mathit{trans}(\mathcal{T})} B$ iff $\tilde{C}_{\forall\exists} \sqsubseteq \tilde{D}_{\forall\exists}$. Hence, it suffices to show that $\tilde{C}_0 \sqsubseteq \tilde{D}_0$ iff

[2]Note that here 'unfolding' is not used in the sense of [21] but only means to iteratively replace all defined concepts in C and D by their respective definitions. The correct word in the sense of [21] would be 'expanding'.

$\tilde{C}_{\forall\exists} \sqsubseteq \tilde{D}_{\forall\exists}$. As $\mathcal{T}$ is reduced the translated TBox *trans*$(\mathcal{T})$ differs from $\mathcal{T}$ only in the quantors occurring on the right-hand side of concept definitions. This implies $L_A(\tilde{C}_0) = L_A(\tilde{C}_{\forall\exists})$ and $L_A(\tilde{D}_0) = L_A(\tilde{D}_{\forall\exists})$ for all $A \in N_{\text{prim}}$.

($\Leftarrow$) The subsumption $\tilde{C}_{\forall\exists} \sqsubseteq \tilde{D}_{\forall\exists}$ especially implies $L_A(\tilde{C}_{\forall\exists}) \supseteq L_A(\tilde{C}_{\forall\exists})$ for all $A \in N_{\text{prim}}$. As argued above, the equality of the role languages of $\tilde{C}_{\forall\exists}$ and $\tilde{C}_0$, and $\tilde{D}_{\forall\exists}$ and $\tilde{D}_0$, respectively, immediately yields $L_A(\tilde{C}_0) \supseteq L_A(\tilde{C}_0)$ for all $A \in N_{\text{prim}}$. By the characterization of subsumption for $\mathcal{FL}_0$ we thus infer $\tilde{C}_0 \sqsubseteq \tilde{D}_0$.

($\Rightarrow$) As seen above we already know that $L_A(\tilde{C}_{\forall\exists}) \supseteq L_A(\tilde{D}_{\forall\exists})$ for all A. In order to prove $\tilde{C}_{\forall\exists} \sqsubseteq \tilde{D}_{\forall\exists}$ it suffices to show that

$$L_\top(\tilde{C}_{\forall\exists}) \cup \bigcup_{A\in N_{\text{prim}}} L_A(\tilde{C}_{\forall\exists}) \cup \{\varepsilon\} \supseteq L_\top(\tilde{D}_{\forall\exists}).$$

We show that $L_\top(\tilde{D}_0)$, which equals $L_\top(\tilde{D}_{\forall\exists})$, is either empty or contains only the empty word ε. Proof by induction on the cardinality $|\mathcal{T}|$ of $\mathcal{T}$.

- $|\mathcal{T}| = 1$
  Then $\mathcal{T} = \{B \doteq D\}$. Since $\mathcal{T}$ is reduced the inapplicability of the first and third reduction rule guarantees that $\top$ can occur in D only on the topmost role level and hence $L_\top(\tilde{D}_0)$ does not contain a word of lenght greater than 0.

- $|\mathcal{T}| > 1$
  Let $E \doteq F$ denote a concept definition in $\mathcal{T}$ where F does not contain defined concepts. Such a definition exists because of the acyclicity of $\mathcal{T}$. If $E = B$ and hence $F = D$ then it suffices to prove the claim w.r.t. the TBox $\{B \doteq D\}$ so that the first case, i.e., $|\mathcal{T}| = 1$, applies.

  If $E \neq B$ then the definition of E is or is not required for the unfolding of D. If it is not then we may prove the claim w.r.t. the TBox $\mathcal{T}\setminus\{E \doteq F\}$ for which the claim holds by induction hypothesis.

  If E is required for the unfolding then the fact that F contains no defined concepts implies that D can be unfolded to a concept D' in which E is the only remaining defined concept. By induction hypothesis (for $\mathcal{T} \setminus \{E \doteq F\}$, treating E as an atomic concept) we know that $\top$ does not occur in D'. The fact that E occurs in D' implies that $F \neq \top$ in the reduced TBox $\mathcal{T}$, because otherwise the second reduction rule would be applicable to $\mathcal{T}$. Moreover, $\top$ does not occur in F because otherwise the first or the third reduction rule would be applicable. Hence, replacing E by F in D', i.e., unfolding D completely produces a concept description in which $\top$ does not occur.

■

It is shown in [21] that subsumption in $\mathcal{FL}_0$ w.r.t acyclic TBoxes (containing only definitions) is co-NP hard. By means of the above reduction, we can immediately infer the following.

Corollary 25 *Deciding subsumption in $\mathcal{L}_{\forall\exists}$ w.r.t. acyclic TBoxes without GCIs or SRIs is co-NP-hard.*

5 Conclusion

We have seen how subsumption in $\mathcal{ELH}$ w.r.t. general TBoxes can be decided in polynomial time. Moreover, it has been shown that the polynomial upper bound does not reach as far as to the DLs $\mathcal{ELN}$, $\mathcal{ELU}$, and $\mathcal{EL}_{\forall\exists}$, where the subsumption problem is co-NP hard even without GCIs.

The attractive complexity and relatively simple structure of the subsumption algorithm naturally motivates the question of how efficient an implementation might be. Even more so, since (i) real-world terminologies such as SNOMED exist which can be classified by our algorithm, and, (ii) the DL systems usually employed for general terminologies implement—highly optimized—EXPTIME algorithms [16, 13].

Two directions of future investigation suggest themselves: firstly, to study other inference problems w.r.t. general $\mathcal{ELH}$-TBoxes; and secondly, to extend $\mathcal{ELH}$ by additional constructors.

Regarding the first direction, the instance problem might be interesting. The problem is solvable in polynomial time w.r.t. cyclic $\mathcal{EL}$ terminologies with descriptive semantics [5]. As we have just seen that the subsumption problem remains polynomial under the transition from cyclic to general terminologies, the same might hold for the instance problem.

For the second direction, desirable constructors might be features, inverse roles, or probably even complex role inclusion axioms. This (far reaching) extension would enable one to reason over the representation language underlying the GALEN [23] terminology. While the polynomial upper bound would undoubtedly be exceeded by this extension, still a complexity better than EXPTIME might be feasible.

References

[1] F. Baader, J. Hladik, C. Lutz, and F. Wolter, 'From tableaux to automata for description logics', in *Proceedings of the 10th International Conference on Logic for Programming, Artificial Intelligence, and Reasoning (LPAR 2003)*, eds., Moshe Vardi and Andrei Voronkov, volume 2850 of *Lecture Notes in Computer Science*, pp. 1–32. Springer, (2003).

[2] F. Baader, R. Küsters, A. Borgida, and D. McGuinness, 'Matching in description logics', *Journal of Logic and Computation*, **9**(3), 411–447, (1999).

[3] F. Baader and P. Narendran, 'Unification of concept terms in description logics', in *Proceedings of the 13th European Conference on Artificial Intelligence (ECAI-98)*, ed., H. Prade, pp. 331–335. John Wiley & Sons Ltd, (1998).

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

[5] 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*, pp. 64–78, Hamburg, Germany, (2003). Springer-Verlag.

[6] Franz Baader, 'Terminological cycles in a description logic with existential restrictions', in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, eds., Georg Gottlob and Toby Walsh, pp. 325–330. Morgan Kaufmann, (2003).

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

[8] R. Cote, D. Rothwell, J. Palotay, R. Beckett, and L. Brochu, 'The systematized nomenclature of human and veterinary medicine', Technical report, SNOMED International, Northfield, IL, (1993).

[9] Robert Dionne, Eric Mays, and Frank J. Oles, 'The equivalence of model-theoretic and structural subsumption in description logics', in

Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence, ed., Ruzena Bajcsy, pp. 710–716, San Mateo, California, (1993). Morgan Kaufmann.

[10] F.M. Donini, 'Complexity of reasoning', in *The Description Logic Handbook: Theory, Implementation, and Applications*, eds., Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, 96–136, Cambridge University Press, (2003).

[11] M. R. Garey and D. S. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, W. H. Freeman and Company, New York, New York, 1979.

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

[13] Volker Haarslev and Ralf Möller, 'RACER system description', *Lecture Notes in Computer Science*, **2083**, 701–712, (2001).

[14] 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).

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

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

[17] 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).

[18] C. Lutz, 'Complexity of terminological reasoning revisited', in *Proceedings of the 6th International Conference on Logic for Programming and Automated Reasoning LPAR'99*, Lecture Notes in Artificial Intelligence, pp. 181–200. Springer-Verlag, (September 6 – 10, 1999).

[19] D. Nardi and R.J. Brachmann, 'An introduction to description logics', in *The Description Logic Handbook: Theory, Implementation, and Applications*, eds., Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, 1–40, Cambridge University Press, (2003).

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

[21] Bernhard Nebel, 'Terminological reasoning is inherently intractable', *Artificial Intelligence*, **43**, 235–249, (1990).

[22] A. Rector, 'Medical informatics', in *The Description Logic Handbook: Theory, Implementation, and Applications*, eds., Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, 406–426, Cambridge University Press, (2003).

[23] 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).

[24] 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*, pp. 414–418, (1993).

[25] K. Spackman, 'Normal forms for description logic expressions of clinical concepts in SNOMED RT', *Journal of the American Medical Informatics Association*, (Symposium Supplement), (2001).