Polynomial Time Reasoning in a Description Logic with Existential Restrictions, GCI Axioms, and—What Else?

Sebastian Brandt[1]

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. One implication of the first result is that reasoning over the widely used medical terminology SNOMED is possible in polynomial time.

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 [17]. 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 [22], GCIs are used especially for two purposes [20]:

- 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 $\ldots \sqsubseteq \exists$has_status.pathological.
- to bridge levels of granularity and add implied meaning to concepts. A classical example [13] is to use a GCI like

  $$\text{ulcer} \sqcap \exists \text{has\_loc.stomach} \sqsubseteq \text{ulcer} \sqcap \exists \text{has\_loc.}(\text{lining} \sqcap \exists \text{is\_part\_of.stomach})$$

  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 [22]. 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 [7] corresponds to an $\mathcal{EL}$-TBox [23]. The representation language underlying the medical terminology GALEN [22] in which GCIs are used extensively, similarly can be represented by a general $\mathcal{EL}$ TBox, 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., $\mathcal{ALCNR}$ [6] and $\mathcal{SHIQ}$ [14], in which deciding subsumption of concepts w.r.t. general TBoxes is EXPTIME hard. Fewer results exist for DLs below $\mathcal{ALC}$. In [11] 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 [9]. 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 [16], 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 [4]. 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.

The paper is organized as follows. Section 2 introduces basic notions essential to study the DLs under consideration. We show in Section 3 that admitting both GCIs and simple role inclusion axioms at the same time preserves the upper bound for subsumption. We also show that the standard technique to decide subsumption in $\mathcal{EL}$ w.r.t. general TBoxes, a tableaux algorithm for $\mathcal{ALC}$, does not guarantee this upper bound. In Section 4 we show that subsumption becomes co-NP hard when $\mathcal{EL}$ is extended by one of the constructors number restriction, disjunction, and allsome. All details and full proofs of the results can be found in our technical report [5].

[1] Theoretical Computer Science, TU Dresden, D-01062 Dresden, Germany. Email: brandt@tcs.inf.tu-dresden.de. Supported by the DFG under Grant BA 1122/4-3.

2 DESCRIPTION LOGICS

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.

Table 1. Syntax and semantics of concept descriptions.

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

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 $r \in N_{\text{role}}$ 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 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 $A \doteq D$ (called *definition*) or $r \sqsubseteq s$ (called *simple role inclusion axiom* (SRI)), where C and D are concept descriptions defined in $\mathcal{L}$, $A \in N_{\text{con}}$, 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 $A \doteq D$ it holds that $A^{\mathcal{I}} = D^{\mathcal{I}}$, 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* [18]. 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 TBox shown in Figure 1, representing in an extremely simplified fashion a part of a medical terminology.

The TBox contains four GCIs and one SRI, stating, e.g., that Pericardium is tissue contained in the heart and that a disease located in a component of the heart is a heart disease and requires treatment.

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

Figure 1. Example $\mathcal{ELH}$-terminology

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.

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 [2] as an example, one can show that this approach in general does not bear fruit. In [5] we give an example $\mathcal{EL}$-TBox for which the tableaux algorithm takes exponentially many steps in the worst case.

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.

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 $\top$.

Such a normal form can easily be computed in polynomial time and does not increase the size of the TBox more than polynomially. An appropriate normalization function is defined in [5]. 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 3 *(Implication set) Let $\mathcal{T}$ denote a normalized $\mathcal{ELH}$-TBox $\mathcal{T}$ over $N_{\text{con}}^{\mathcal{T}}$ and $N_{\text{role}}^{\mathcal{T}}$. 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 $A \in S_*(B)$.

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

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 4 *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 $A \sqsubseteq_{\mathcal{T}} B$.*

Due to space limitations , we can only give a proof sketch. The full proof is shown in [5]. Claim (i) is trivial. 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 $i > 0$, 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. For the canonical model $\mathcal{I}$ 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**, see [5].

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 $S_i(A) \subseteq N_{\text{con}}$ and $S_i(r) \subseteq N_{\text{role}}$. 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 5 *Subsumption in $\mathcal{ELH}$ w.r.t. general TBoxes 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 representation perspective, particularly useful constructors might be number restrictions ($\leq n\, r$) and ($\geq n\, r$), and disjunction ($\sqcup$). The DL K-REP [8] 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 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 6 *(*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 [10, 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 top-level 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 7 *(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*

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

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:

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

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. Note that one instance of f^P can be computed easily in polynomial time.

Lemma 8 *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, P has a solution iff C^P is satisfiable.*

The concept descriptions in $\mathcal{C}^P$ correspond to binary numbers from 0 to $\Sigma_{u \in U} s(u) =: w$, the overall weight of all $u \in U$. The injectivity of f^P over $\mathcal{C}^P$ enforces that $f^P(u,i) \sqcap f^P(v,j)$ is inconsistent iff $u \neq v$ or $i \neq j$, implying at least w vertices on role level 2 in every model of C^P. On top-level, C^P requires one existential successor for every $u \in U$. Hence, C^P is satisfiable iff these $|U|$ r-successors, which do not have to be distinct in a model, can be represented by k r-successors of the root vertex such that each successor has at most b distinct r-successors. Hence, satisfiability is equivalent to P being solvable. For the full proof, see [5]. As satisfiability of $\mathcal{ELN}$-concepts can be reduced to non-subsumption, i.e., C satisfiable iff $C \not\sqsubseteq (\leq 0\, r) \sqcap (\geq 1\, r)$, we immediately obtain the hardness results for subsumption.

Corollary 9 *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 10 *(*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 [10, 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 $C \not\sqsubseteq \text{nnf}(\neg D)$, where $\text{nnf}(\neg D)$ denotes the negation normal form of $\neg D$. Note that $\text{nnf}(\neg D)$ is in fact an $\mathcal{ELU}$-concept description. (See [5] for details.)

Corollary 11 *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 12 *(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:*

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

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

Note that the above definition is correct only in the sense that all subsumption relations are preserved. While a model of $trans(\mathcal{T})$ can always be shown to be model of $\mathcal{T}$, the reverse 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 [19] or *concept centered normal form* in [1]. Given a concept description C, the idea is to exploit the above equivalence from left to right until conjunction in C occurs only on top-level, 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 $\forall\exists r.C \sqcap \forall\exists r.D$. Therefore, the following definition is justified.

Definition 13 *(Role languages) Let C be an $\mathcal{L}_{\forall\exists}$-concept description. Then, for $A \in N_{\text{prim}} \cup \{\top\}$ the formal language $L_A(C) \subset N^*_{\text{prim}}$ is inductively defined by:*

$$\begin{aligned} L_A(B) &:= \{\varepsilon \mid A = B\} \\ L_A(C \sqcap D) &:= L_A(C) \cup L_A(D) \\ L_A(\forall\exists r.C) &:= \{r\} \cdot L_A(C), \end{aligned}$$

where B is an arbitrary concept name $B \in N_{\text{prim}}$ or $B = \top$.

The language $L_A(C)$ contains all words $r_1 \ldots r_n$ over N_{role} with $C \sqsubseteq \forall\exists r_1.\cdots\forall\exists r_n.A$. This fact can be exploited for a role-language characterization of subsumption of $\mathcal{L}_{\forall\exists}$-concept descriptions w.r.t. the empty TBox.

Lemma 14 *Let C, D be $\mathcal{L}_{\forall\exists}$-concept descriptions over $N_{\rm prim}$ and $N_{\rm role}$. Then, $C \sqsubseteq D$ iff*

1. *$L_A(C) \supseteq L_A(D)$ for all $A \in N_{\rm role}$; and*
2. *$L_\top(C) \cup \bigcup_{A \in N_{\rm prim}} L_A(C) \cup \{\varepsilon\} \supseteq L_\top(D)$.*

To show ($\Rightarrow$) we assume that one of the subset relations is violated and construct an appropriate model where the subsumption $C \sqsubseteq D$ does not hold. The reverse direction ($\Leftarrow$) utilizes the equivalence $\forall\exists r.(C \sqcap D) \equiv \forall\exists r.C \sqcap \forall\exists r.D$ to rewrite C syntactically to the form $C = D \sqcap R$, implying the subsumption. (See [5] for details.) The above characterization of subsumption allows a straightforward proof of correctness of the translation from $\mathcal{FL}_0$ to $\mathcal{L}_{\forall\exists}$.

Lemma 15 *Let $\mathcal{T}$ be an acyclic reduced $\mathcal{FL}_0$-TBox over $N_{\rm con}$, and $N_{\rm role}$. Let $A, B \in N_{\rm def}$. Then, $A \sqsubseteq_{\mathcal{T}} B$ iff $A \sqsubseteq_{trans(\mathcal{T})} B$.*

Denote by $\tilde{A}, \tilde{B}$ the descriptions of A, B fully expanded w.r.t. $\mathcal{T}$, and analogously by $\tilde{A}_{\rm tr}, \tilde{B}_{\rm tr}$ those expanded w.r.t. $trans(T)$. As $\mathcal{T}$ and $trans(\mathcal{T})$ have the same structure, $L_C(\tilde{A})$ equals $L_C(\tilde{A}_{\rm tr})$ for every $C \in N_{\rm prim}$ (and analogously for B). Condition 1 of Lemma 14 characterizes subsumption of $\mathcal{FL}_0$-concept descriptions [19], implying for the proof direction ($\Rightarrow$) that it suffices to show Condition 2. Condition 2 holds because differences w.r.t. the top concept semantically 'vanish' under translation from $\mathcal{L}_{\forall\exists}$ to $\mathcal{FL}_0$, where always $\forall r.\top \equiv \top$. For the reverse direction ($\Leftarrow$) we show by induction on the number of definitions in $\mathcal{T}$ that the role language $L_\top(\tilde{B})$ is either empty or equals $\{\varepsilon\}$, satisfying Lemma 14. (See [5] for details.)

Corollary 16 *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 [15, 12].

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 [3]. 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 [21] terminology. While the polynomial upper bound would undoubtedly be exceeded by this extension, still a complexity better than EXPTIME might be feasible.

ACKNOWLEDGEMENTS

We wish to express our thanks to Carsten Lutz for a multitude of useful remarks and ideas that have greatly influenced this work.

REFERENCES

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

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

[3] F. 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 *LNAI*, pp. 64–78, (2003).

[4] F. Baader, 'Terminological cycles in a description logic with existential restrictions', in *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pp. 325–330. Morgan Kaufmann, (2003).

[5] S. Brandt, 'Reasoning in $\mathcal{ELH}$ w.r.t. General Concept Inclusion Axioms', Technical report, (2004). See http://lat.inf.tu-dresden.de /research/reports.html.

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

[7] 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).

[8] R. Dionne, E. Mays, and F. 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., R. Bajcsy, pp. 710–716, Morgan Kaufmann, (1993).

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

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

[11] R. Givan, D. A. McAllester, C. Witty, and D. Kozen, 'Tarskian set constraints', *Information and Computation*, **174(2)**, 105–131, (2002).

[12] V. Haarslev and R. Möller, 'RACER system description', *LNCS*, **2083**, 701–712, (2001).

[13] I. Horrocks, A. L. Rector, and C. A. Goble, 'A description logic based schema for the classification of medical data', in *Knowledge Representation Meets Databases*, (1996).

[14] I. Horrocks, U. Sattler, and S. Tobies, 'Practical reasoning for expressive description logics', in *Proceedings of the 6th International Conference on Logic for Programming and Automated Reasoning (LPAR'99)*, number 1705 in LNAI, pp. 161–180. Springer-Verlag, (1999).

[15] I. Horrocks, 'Using an expressive description logic: FaCT or fiction?', in *KR'98: Principles of Knowledge Representation and Reasoning*, pp. 636–645, Morgan Kaufmann, (1998).

[16] Y. Kazakov and H. 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).

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

[18] B. Nebel, 'Terminological cycles: Semantics and computational properties', in *Principles of Semantic Networks: Explorations in the Representation of Knowledge*, 331–361, Morgan Kaufmann Publishers, (1991).

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

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

[21] 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).

[22] 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).

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