Subsumption in $\mathcal{EL}$ w.r.t. hybrid TBoxes

Sebastian Brandt and Jörg Model

Theoretical Computer Science,
TU Dresden,
D-01062 Dresden, Germany.
lastname@tcs.inf.tu-dresden.de

Abstract. In the area of Description Logic (DL) based knowledge representation, two desirable features of DL systems have as yet been incompatible: firstly, the support of general TBoxes containing general concept inclusion (GCI) axioms, and secondly, non-standard inference services facilitating knowledge engineering tasks, such as build-up and maintenance of terminologies (TBoxes).
In order to make non-standard inferences available without sacrificing the convenience of GCIs, the present paper proposes *hybrid TBoxes* consisting of a pair of a general TBox $\mathcal{F}$ interpreted by descriptive semantics, and a (possibly) cyclic TBox $\mathcal{T}$ interpreted by fixpoint semantics. $\mathcal{F}$ serves as a foundation of $\mathcal{T}$ in the sense that the GCIs in $\mathcal{F}$ define relationships between concepts used as atomic concept names in the definitions in $\mathcal{T}$. Our main technical result is a polynomial time subsumption algorithm for hybrid $\mathcal{EL}$-TBoxes based on a polynomial reduction to subsumption w.r.t. cyclic $\mathcal{EL}$-TBoxes with fixpoint semantics. By virtue of this reduction, all non-standard inferences already available for cyclic $\mathcal{EL}$-TBoxes become available for hybrid ones.

1 Motivation

In the area of Description Logic (DL) based knowledge representation (KR), intensional knowledge of a given domain is represented by a terminology (TBox) that defines general properties of concepts relevant to the domain [1]. In its simplest form, a TBox comprises *definitions* of the form $A \equiv C$ by which a *concept name* A is assigned to a *concept description* C. Concept descriptions are terms built from atomic concepts by means of a set of constructors provided by the DL under consideration. In the present case, we are mostly concerned with the DL $\mathcal{EL}$ which provides conjunction ($\sqcap$) and existential restriction ($\exists r.C$).

General TBoxes additionally allow for *general concept inclusion (GCI)* axioms of the form $C \sqsubseteq D$, where both C and D may be complex concept descriptions. GCIs define implications ("D holds whenever C holds") relevant to the terminology as a whole. The utility of GCIs for practical KR applications has been examined in depth; see, e.g., [2–4]. Apart from constraining terminologies further without explicitly changing all its definitions, using GCIs can lead to smaller, more readable TBoxes, and can facilitate the re-use of data in applications of different levels of detail. As a consequence, GCIs are supported by most modern DL systems, such as FaCT [5] and Racer [6].

TBoxes are interpreted w.r.t. a model-theoretic *semantics* which allows to reason over the terminology in a formally well-defined way. A model $\mathcal{I}$ satisfies a definition $A \equiv C$ iff the extensions of A and C in $\mathcal{I}$ are equal. A GCI $C \sqsubseteq D$ is satisfied by $\mathcal{I}$ iff the extension of C is a subset of the one of D. $\mathcal{I}$ is a model of a TBox $\mathcal{T}$ iff all definitions and GCIs in $\mathcal{T}$ are satisfied. This semantics for TBoxes is usually called *descriptive semantics* [7]. In contrast, *greatest fixpoint semantics* only considers models that interpret concepts as large as possible.

One of the most important reasoning services provided by DL systems is computing the subsumption hierarchy. A concept A is *subsumed* by a concept B w.r.t. a TBox $\mathcal{T}$ iff the extension of A is a subset of that of B in every model of $\mathcal{T}$. Before DL systems can be used to reason over terminologies, however, the relevant TBoxes must be built-up and maintained. In order to support these knowledge engineering tasks, additional so-called 'non-standard' inference services have been proposed, most notably *least-common subsumer (lcs)* [8–10], *most specific concept (msc)* [10], and *matching* [11–13]. It has been argued in [14] that the lcs and msc facilitate the build-up of DL terminologies in a 'bottom-up' fashion suitable for domain experts with limited KR background. 'Bottom-up' fashion means to begin by selecting a set of example instances and use them to construct a new concept description intended to represent them. Matching, on the other hand, can be used as a means of querying TBoxes for concepts of a certain structure [15]. This can be utilized to construct new concepts by retrieving and modifying structurally similar ones in the TBox.

The practical utility of GCIs and non-standard inferences motivates the question for a DL in which both can be provided. The problem encountered here relates to the appropriate choice of semantics. General TBoxes need to be interpreted w.r.t. *descriptive* semantics. For this kind of semantics, however, it has been shown in [16] that lcs and msc need not always exist, even w.r.t. cyclic $\mathcal{EL}$-TBoxes. This result carries over to general $\mathcal{EL}$-TBoxes and any extension of $\mathcal{EL}$. The same holds for matching which relies on the lcs. On the other hand, lcs and msc are available for cyclic $\mathcal{EL}$-TBoxes interpreted by *fixpoint semantics* [16].

In order to provide lcs and msc without sacrificing the convenience of GCIs, the present paper proposes *hybrid TBoxes*. A hybrid $\mathcal{EL}$-TBox is a pair $(\mathcal{F}, \mathcal{T})$ of a general TBox $\mathcal{F}$ ('foundation') and a possibly cyclic TBox $\mathcal{T}$ ('terminology') defined over the same set of atomic concepts and roles. $\mathcal{F}$ serves as a foundation of $\mathcal{T}$ in that the GCIs in $\mathcal{F}$ define relationships between concepts used as atomic concept names in the definitions in $\mathcal{T}$. Hence, $\mathcal{F}$ lays a foundation of general implications constraining $\mathcal{T}$. Models of $(\mathcal{F}, \mathcal{T})$ are greatest fixpoint models of $\mathcal{T}$ that respect all GCIs in $\mathcal{F}$. Hence, the foundation is interpreted by descriptive semantics while the terminology is interpreted by greatest fixpoint semantics. Note that hybrid $\mathcal{EL}$-TBoxes cannot be reduced to ordinary general $\mathcal{EL}$-TBoxes.

Having introduced hybrid TBoxes, the main purpose of the present paper is to show that subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes can be decided by a polynomial reduction to cyclic $\mathcal{EL}$-TBoxes for which a polynomial time decision algorithm exists [17]. This yields a polynomial time subsumption algorithm for hybrid $\mathcal{EL}$-TBoxes. An implication of this reduction is that non-standard inferences available for cyclic $\mathcal{EL}$-TBoxes can directly be utilized for hybrid $\mathcal{EL}$-TBoxes. In this sense, our initial goal of providing non-standard inferences in the presence of GCIs is met.

Another application for hybrid TBoxes might be DL systems supporting users with limited KR background. By restricting the view to the definitions in the terminology while hiding the GCIs in the foundation, the system could provide a simplified version of its knowledge base while preserving correct inferences.

The present paper is organized as follows. Basic definitions related to cyclic $\mathcal{EL}$-TBoxes are introduced in Section 2 while Section 3 formally introduces hybrid $\mathcal{EL}$-TBoxes. In Section 3.2 we show how subsumption w.r.t. hybrid TBoxes can be reduced to subsumption w.r.t. cyclic $\mathcal{EL}$-TBoxes interpreted by greatest fixpoint semantics.

2 Cyclic $\mathcal{EL}$-TBoxes

We begin by formally introducing syntax and semantics of cyclic $\mathcal{EL}$ TBoxes. In fact, we will consider two different semantics first introduced by Nebel [7], namely descriptive semantics and greatest[1] fixpoint semantics. Most of our preliminary Sections 2.2 and 2.3 recall basic definitions and results from [17].

2.1 Syntax and (descriptive) semantics

Concept descriptions are inductively defined with the help of a set of concept *constructors*, starting with disjoint sets N_{prim} and N_{def} of *primitive concept names* and *defined concept names*, respectively, and a set N_{role} of *role names*. In this paper, we consider the small DL $\mathcal{EL}$ which provides the concept constructors top-concept ($\top$), conjunction ($C \sqcap D$), and existential restrictions ($\exists r.C$).

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 $P \in N_{\text{prim}} \cup N_{\text{def}}$ 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 follows.

$$\begin{aligned}
\top^{\mathcal{I}} &:= \Delta^{\mathcal{I}} \\
(C \sqcap D)^{\mathcal{I}} &:= C^{\mathcal{I}} \cap D^{\mathcal{I}} \\
(\exists r.C)^{\mathcal{I}} &:= \{x \in \Delta^{\mathcal{I}} \mid \exists y\colon (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}
\end{aligned}$$

Definition 1. *An $\mathcal{EL}$-terminology (called $\mathcal{EL}$-TBox) is a finite set $\mathcal{T}$ of* definitions *of the form $A \equiv C$, where $A \in N_{\text{def}}$ is a concept name and C is an $\mathcal{EL}$-concept description over N_{prim}, N_{def}, and N_{role}. For every such definition, A is called* defined in $\mathcal{T}$ *and may occur on the left-hand side of no other definition in $\mathcal{T}$. Note that cyclic definitions are allowed, i.e., every defined concept may occur on the right-hand side of every definition.*

[1] It has been argued in [17] that least fixpoint semantics is not interesting for cyclic $\mathcal{EL}$-TBoxes because cycles are always interpreted by the empty set.

The size *of $\mathcal{T}$ is defined as the sum of the sizes of all definitions in $\mathcal{T}$. Denote by $N_{\text{prim}}^{\mathcal{T}}$, $N_{\text{def}}^{\mathcal{T}}$, and $N_{\text{role}}^{\mathcal{T}}$ the set of all primitive concepts, defined concept names, and role names, respectively, occurring in $\mathcal{T}$.*

The semantics of a cyclic $\mathcal{EL}$-TBox can now be defined as follows.

Definition 2. *An interpretation $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is a* model *of $\mathcal{T}$ ($\mathcal{I} \models \mathcal{T}$) iff $A^{\mathcal{I}} = C^{\mathcal{I}}$ for all definitions $A \equiv C \in \mathcal{T}$.*

This semantics has been introduced as *descriptive semantics* by Nebel [7]. One of the most basic inference services provided by DL systems is computing the subsumption hierarchy. Formally, (descriptive) subsumption is defined as follows.

Definition 3. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and let $A, B \in N_{\text{def}}^{\mathcal{T}}$. Then, A is subsumed by B w.r.t. descriptive semantics ($A \sqsubseteq_{\mathcal{T}} B$) iff $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ holds for all models $\mathcal{I}$ of $\mathcal{T}$.*

The other type of semantics relevant for us is introduced in the following section.

2.2 Greatest fixpoint semantics

In contrast to descriptive semantics, some more formal preliminaries are necessary to define greatest fixpoint (gfp) semantics. Recalling the relevant definitions from [17], we begin by introducing a normal form for cyclic $\mathcal{EL}$-TBoxes.

Definition 4. *An $\mathcal{EL}$-TBox $\mathcal{T}$ is* normalized *iff $A \equiv D \in \mathcal{T}$ implies that D is of the form*

$$P_1 \sqcap \cdots \sqcap P_m \sqcap \exists r_1.B_1 \sqcap \ldots \exists r_\ell.B_\ell,$$

where for $m, \ell \geq 0$, $P_1, \ldots, P_m \in N_{\text{prim}}$ and $B_1, \ldots, B_\ell \in N_{\text{def}}$. If $m = \ell = 0$ then $D = \top$.

In order to refer to the definition of a defined concept more conveniently, the following notation is introduced for normalized $\mathcal{EL}$-TBoxes.

Definition 5. *For a normalized $\mathcal{EL}$-TBox $\mathcal{T}$ and every $A \in N_{\text{def}}^{\mathcal{T}}$, let*

$$\text{def}_{\mathcal{T}}(A) := \{P_1, \ldots, P_m\} \cup \{\exists r_1.B_1, \ldots, \exists r_\ell.B_\ell\}$$

iff A is defined in $\mathcal{T}$ by $A \equiv P_1 \sqcap \cdots \sqcap P_m \sqcap \exists r_1.B_1 \sqcap \ldots \exists r_\ell.B_\ell$.

A gfp-model for a given $\mathcal{EL}$-TBox $\mathcal{T}$ is obtained in two steps. In the first step, only the primitive concepts and roles occurring in $\mathcal{T}$ are interpreted. The second step comprises an iteration by which the interpretation of the defined names in $\mathcal{T}$ is changed until a fixpoint is reached. The following definition formalizes the first step.

Definition 6. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox over N_{prim}, N_{role}, and N_{def}. A* primitive interpretation $(\Delta^{\mathcal{J}}, \cdot^{\mathcal{J}})$ *of $\mathcal{T}$ interprets all primitive concepts $P \in N_{\text{prim}}$ by subsets of $\Delta^{\mathcal{J}}$ and all roles $r \in N_{\text{role}}$ by binary relations on $\Delta^{\mathcal{J}}$. An Interpretation $(\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$ is* based on $\mathcal{J}$ *iff $\Delta^{\mathcal{J}} = \Delta^{\mathcal{I}}$ and $\cdot^{\mathcal{J}}$ and $\cdot^{\mathcal{I}}$ coincide on N_{role} and N_{prim}. The set of all initerpretations based on $\mathcal{J}$ is denoted by*

$$\text{Int}(\mathcal{J}) := \{\mathcal{I} \mid \mathcal{I} \textit{ is an interpretation based on } \mathcal{J}\}.$$

On $\text{Int}(\mathcal{J})$, a binary relation $\preceq_{\mathcal{J}}$ is defined for all $\mathcal{I}_1, \mathcal{I}_2 \in \text{Int}(\mathcal{J})$ by

$$\mathcal{I}_1 \preceq_{\mathcal{J}} \mathcal{I}_2 \quad \textit{iff} \quad A^{\mathcal{I}_1} \subseteq A^{\mathcal{I}_2} \textit{ for all } A \in N_{\text{def}}^{\mathcal{T}}.$$

Primitive interpretations do not interpret defined concepts from N_{def}. It is easy to see that $\preceq_{\mathcal{J}}$ is a complete lattice on $\text{Int}(\mathcal{J})$, so that every subset of $\text{Int}(\mathcal{J})$ has a least upper bound (lub) and a greatest lower bound (glb). Hence, by Tarski's fixpoint theorem [18], every monotonic function on $\text{Int}(\mathcal{J})$ has a fixpoint. In particular, this applies to the function $O_{\mathcal{T},\mathcal{J}}$ to be introduced next.

Definition 7. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox over N_{prim}, N_{role}, and N_{def}, and $\mathcal{J}$ a primitive interpretation of N_{prim} and N_{role}. Then $O_{\mathcal{T},\mathcal{J}}$ is defined as follows.*

$$\begin{aligned} O_{\mathcal{T},\mathcal{J}} \colon & \text{Int}(\mathcal{J}) \to \text{Int}(\mathcal{J}) \\ & \mathcal{I}_1 \mapsto \mathcal{I}_2 \textit{ iff } A^{\mathcal{I}_2} = C^{\mathcal{I}_1} \textit{ for all } A \equiv C \in \mathcal{T}. \end{aligned}$$

As shown in [17], $O_{\mathcal{T},\mathcal{J}}$ is in fact monotonous and can be used as a fixpoint operator on $\text{Int}(\mathcal{J})$. As a result, we obtain the following proposition.

Proposition 1. *Let $\mathcal{I}$ be an interpretation based on the primitive interpretation $\mathcal{J}$. Then $\mathcal{I}$ is a fixpoint of $O_{\mathcal{T},\mathcal{J}}$ iff $\mathcal{I}$ is a model of $\mathcal{T}$.*

With this, the general notion of fixpoint models for $\mathcal{EL}$-TBoxes can be defined as follows.

Definition 8. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox. The model $\mathcal{I}$ of $\mathcal{T}$ is called* gfp-model *iff there is a primitive interpretation $\mathcal{J}$ such that $\mathcal{I} \in \text{Int}(\mathcal{J})$ is the greatest fixpoint of $O_{\mathcal{T},\mathcal{J}}$.* Greatest fixpoint semantics *considers only gfp-models as admissible models.*

As $(\text{Int}(\mathcal{J}), \preceq_{\mathcal{J}})$ is a complete lattice, the gfp-model is uniquely determined for a given TBox $\mathcal{T}$ and a primitive interpretation $\mathcal{J}$. This allows us to refer to *the* gfp-model $\text{gfp}(\mathcal{T}, \mathcal{J})$ for any given $\mathcal{T}$ and $\mathcal{J}$.

In order to show how the gfp-model $\text{gfp}(\mathcal{T}, \mathcal{J})$ can be obtained, we need to introduce the iteration of $O_{\mathcal{T},\mathcal{J}}$ over *ordinals*.

Definition 9. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox over N_{prim}, N_{role}, and N_{def} and $\mathcal{J}$ a primitive interpretation of N_{prim} and N_{role}. Define $\mathcal{I}^{\text{top}} \in \text{Int}(J)$ by $\mathcal{I}_{\mathcal{J}}^{\text{top}}(A) := \Delta^{\mathcal{J}}$ for every $A \in N_{\text{def}}$. For every ordinal α, define*

– $\mathcal{I}_{\mathcal{T},\mathcal{J}}^{\downarrow\alpha} := \mathcal{I}_{\mathcal{J}}^{\text{top}}$ *if $\alpha = 0$;*

– $\mathcal{I}^{\downarrow\alpha+1}_{\mathcal{T},\mathcal{J}} := O_{\mathcal{T},\mathcal{J}}(\mathcal{I}^{\downarrow\alpha})$;
– $\mathcal{I}^{\downarrow\alpha}_{\mathcal{T},\mathcal{J}} := \mathrm{glb}(\{\mathcal{I}^{\downarrow\beta} \mid \beta < \alpha\})$ *if* α *is a limit ordinal.*

The following corollary now shows that computing $\mathrm{gfp}(\mathcal{T},\mathcal{J})$ is equivalent to computing $\mathcal{I}^{\downarrow\alpha}_{\mathcal{T},\mathcal{J}}$, given an appropriate ordinal α.

Corollary 1. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox over N_{prim}, N_{role}, and N_{def}. Let $\mathcal{J}$ be a primitive interpretation of N_{prim} and N_{role}. Then there exists an ordinal α such that* $\mathrm{gfp}(\mathcal{T},\mathcal{J}) = \mathcal{I}^{\downarrow\alpha}_{\mathcal{T},\mathcal{J}}$.

Note that if α is a limit ordinal then $\mathcal{I}^{\downarrow\alpha}_{\mathcal{T},\mathcal{J}}$ equals $\bigcap_{\beta<\alpha} \mathcal{I}^{\downarrow\beta}_{\mathcal{T},\mathcal{J}}$. With this preparation, we are ready to introduce gfp-subsumption.

Definition 10. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and let $A, B \in N^{\mathcal{T}}_{\mathrm{def}}$. Then, A is subsumed by B w.r.t. gfp-semantics ($A \sqsubseteq_{\mathrm{gfp},\mathcal{T}} B$) iff $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ holds for all gfp-models $\mathcal{I}$ of $\mathcal{T}$.*

Note that descriptive semantics considers a superset of the set of gfp-models, implying that descriptive subsumption entails gfp-subsumption. Hence, all subsumption relations w.r.t. $\sqsubseteq_{\mathcal{T}}$ also hold w.r.t. $\sqsubseteq_{\mathrm{gfp},\mathcal{T}}$. The question of how to *decide* subsumption w.r.t. gfp-semantics is addressed in the following section.

2.3 Deciding subsumption w.r.t. cyclic $\mathcal{EL}$-TBoxes with descriptive semantics

As in the previous section, we begin by recalling some definitions from [17] elementary for the decision procedure for gfp-subsumption.

Definition 11. *An $\mathcal{EL}$-*description graph *is a graph $\mathcal{G} = (V, E, L)$ where*

– *V is a set of nodes;*
– *$E \subseteq V \times N_{\mathrm{role}} \times V$ is a set of edges labeled by role names;*
– *$L\colon V \to 2^{N_{\mathrm{prim}}}$ is a function that labels nodes with sets of primitive concepts.*

Description graphs can be used to represent TBoxes and primitive interpretations. The description graph of a TBox is defined as follows.

Definition 12. *Let $\mathcal{T}$ be a normalized $\mathcal{EL}$-TBox. Then the $\mathcal{EL}$-description graph $\mathcal{G}_{\mathcal{T}} = (N^{\mathcal{T}}_{\mathrm{def}}, E_{\mathcal{T}}, L_{\mathcal{T}})$ of $\mathcal{T}$ is defined as follows:*

– *the nodes of $\mathcal{G}_{\mathcal{T}}$ are the defined concepts of $\mathcal{T}$;*
– *if A is defined in $\mathcal{T}$ and*

$$A \equiv P_1 \sqcap \cdots \sqcap P_m \sqcap \exists r_1.B_1 \sqcap \cdots \sqcap \exists r_\ell.B_\ell$$

is its definition then $L_{\mathcal{T}}(A) := \{P_1, \ldots, P_m\}$, and A is the source of the edges $(A, r_1, B_1), \ldots, (A, r_\ell, B_\ell) \in E_{\mathcal{T}}$.

Note that for every $A \in N_{\text{def}}^{\mathcal{T}}$, $L_{\mathcal{T}}(A)$ can be written as $\text{def}_{\mathcal{T}}(A) \cap N_{\text{prim}}^{\mathcal{T}}$. For primitive definitions, we define description graphs in the following way.

Definition 13. *Let $\mathcal{J} = (\Delta^{\mathcal{J}}, \cdot^{\mathcal{J}})$ be a primitive interpretation. Then the $\mathcal{EL}$-description graph $\mathcal{G}_{\mathcal{J}} = (\Delta^{\mathcal{J}}, E_{\mathcal{J}}, L_{\mathcal{J}})$ of $\mathcal{J}$ is defined as follows:*

- *the nodes of $\mathcal{G}_{\mathcal{J}}$ are the elements of $\Delta^{\mathcal{J}}$;*
- *$E_{\mathcal{J}} := \{(x, r, y) \mid (x, y) \in r^{\mathcal{J}}\}$;*
- *$L_{\mathcal{J}}(x) = \{P \in N_{\text{prim}} \mid x \in P^{\mathcal{J}}\}$ for all $x \in \Delta^{\mathcal{J}}$.*

In preparation for the characterization of subsumption we need to introduce simulation relations on description graphs.

Definition 14. *Let $\mathcal{G}_i = (V_i, E_i, L_i)$, $i = 1, 2$, be two $\mathcal{EL}$-description graphs. The binary relation $Z \subseteq V_1 \times V_2$ is a* simulation relation *from $\mathcal{G}_1$ to $\mathcal{G}_2$ ($Z \colon \mathcal{G}_1 \rightleftarrows \mathcal{G}_2$) iff*

(S1) *$(v_1, v_2) \in Z$ implies $L_1(v_1) \subseteq L_2(v_2)$; and*
(S2) *if $(v_1, v_2) \in Z$ and $(v_1, r, v_1') \in E_1$ then there exists a node $v_2' \in V_2$ such that $(v_1', v_2') \in Z$ and $(v_2, r, v_2') \in E_2$.*

It has been shown in [17] that simulation relations can be concatenated in the sense of the following lemma.

Lemma 1. *Let $\mathcal{G}_i := (V_i, E_i, L_i)$, $i = 1, 2, 3$, be $\mathcal{EL}$-description graphs, and let $Z_1 \colon \mathcal{G}_1 \rightleftarrows \mathcal{G}_2$ and $Z_2 \colon \mathcal{G}_2 \rightleftarrows \mathcal{G}_3$. Then $Z_1 \circ Z_2 \colon \mathcal{G}_1 \rightleftarrows \mathcal{G}_3$, where*

$$Z_1 \circ Z_2 := \{(v, v'') \mid \exists v' \in V_2 \colon (v, v') \in Z_1 \wedge (v', v'') \in Z_2\}.$$

One of the main results in [17] is a characterization of gfp-subsumption by simulation relations over description graphs. The following results provide the relevant characterizations.

Proposition 2. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox over N_{prim}, N_{role}, and N_{def} and $A \in N_{\text{def}}^{\mathcal{T}}$. Let J be a primitive interpretation of N_{prim} and N_{role}. Then $x \in A^{\text{gfp}(\mathcal{T},\mathcal{J})}$ iff there is a simulation relation $Z \colon \mathcal{G}_{\mathcal{T}} \rightleftarrows \mathcal{G}_{\mathcal{J}}$ such that $(A, x) \in Z$.*

Theorem 1. *Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and A, B be defined concepts in $\mathcal{T}$. Then $A \sqsubseteq_{\text{gfp},\mathcal{T}} B$ iff there is a simulation relation $Z \colon \mathcal{G}_{\mathcal{T}} \rightleftarrows \mathcal{G}_{\mathcal{T}}$ such that $(B, A) \in Z$.*

Since the description graph of a TBox is of polynomial size in the size of the TBox and since the existence of simulation relations with the required properties can be tested in polynomial time, the following complexity result is obtained [17].

Corollary 2. *Subsumption w.r.t. gfp-semantics in $\mathcal{EL}$ can be decided in polynomial time.*

With this result, the prerequisites for the introduction of hybrid $\mathcal{EL}$-TBoxes are complete.

3 Hybrid $\mathcal{EL}$-TBoxes

In the present section, we start by defining syntax and semantics of hybrid $\mathcal{EL}$-TBoxes formally before showing in Section 3.2 how subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes can be decided in polynomial time.

3.1 Syntax and semantics

The following definition introduces hybrid $\mathcal{EL}$-TBoxes, the central notion of the present paper.

Definition 15. *A* general concept inclusion axiom (GCI) over N_{prim} and N_{role} *is of the form* $C \sqsubseteq D$*, where* C *and* D *are* $\mathcal{EL}$*-concept descriptions over* N_{prim} *and* N_{role}*. A finite set of GCIs over* N_{prim} *and* N_{role} *is called a* general $\mathcal{EL}$-TBox over N_{prim} and N_{role}*. A primitive concept* $P \in N_{\mathrm{prim}}$ *(an existential restriction* $\exists r.P$ *with* $r \in N_{\mathrm{role}}$*)* occurs in $\mathcal{T}$ *iff there is a GCI* $C \sqsubseteq D \in \mathcal{T}$ *such that* P *(*$\exists r.P$*) is a conjunct of* C *or* D*.*

A hybrid $\mathcal{EL}$-TBox *is a pair* $(\mathcal{F}, \mathcal{T})$*, where* $\mathcal{F}$ *is a general* $\mathcal{EL}$*-TBox over* N_{prim} *and* N_{role} *and* $\mathcal{T}$ *is an* $\mathcal{EL}$*-TBox over* N_{prim}*,* N_{role}*, and* N_{def}*.*

Note that our general TBoxes are restricted 'over N_{prim} and N_{role}' to rule out the use of defined concepts from $\mathcal{T}$. Similar to the case of cyclic $\mathcal{EL}$-TBoxes, we introduce a normal form for hybrid $\mathcal{EL}$-TBoxes in order to simplify our solution for the respective subsumption problem.

Definition 16. *Let* $(\mathcal{F}, \mathcal{T})$ *be a hybrid TBox over* N_{prim}*,* N_{role}*, and* N_{def}*. Then,* $(\mathcal{F}, \mathcal{T})$ *is* normalized *iff*

1. *Every GCI in* $\mathcal{F}$ *is of one of the following forms:* $A \sqsubseteq B$*,* $A_1 \sqcap A_2 \sqsubseteq B$*,* $A \sqsubseteq \exists r.B$*, or* $\exists r.A \sqsubseteq B$*, where* $r \in N_{\mathrm{role}}$ *and* $A, A_1, A_2, B \in N_{\mathrm{prim}} \cup \{\top\}$*;*
2. $\mathcal{T}$ *is normalized in the sense of Definition 4; and*
3. *for every primitive concept* P *and for every existential restriction* $\exists r.P$ *occurring in* $\mathcal{F}$*,* $\mathcal{T}$ *contains a definition of the form* $A_P \equiv P$ *and* $A_{\exists r.P} \equiv \exists r.A_P$*, respectively.*

Note that the first two normalization conditions can be satisfied easily for any hybrid TBox $(\mathcal{F}, \mathcal{T})$, see [17]. For the third condition, a conservative extension $\mathcal{T}'$ of $\mathcal{T}$ of size at most the size of $(\mathcal{F}, \mathcal{T})$ can be found such that $(\mathcal{F}, \mathcal{T}')$ is normalized. All subsumption relations between concept names defined in $\mathcal{T}$ remain unchanged.

Example 1. In order to get an impression of how an actual hybrid TBox might look like, consider Figure 1. Shown is an extremely simplified part of a medical terminology[2] represented by a hybrid TBox $(\mathcal{F}, \mathcal{T})$. $\mathcal{T}$ is supposed to define the concepts 'disease of the connective tissue', 'bacterial infection' and 'bacterial

[2] Our example is only supposed to show the features of hybrid $\mathcal{EL}$-TBoxes and in no way claims to be adequate from a Medical KR perspective.

$\mathcal{T}$:	$\mathit{ConnTissDisease} \equiv \mathsf{Disease} \sqcap \exists \mathsf{acts\_on}.\mathsf{ConnTissue}$
	$\mathit{BactInfection} \equiv \mathsf{Infection} \sqcap \exists \mathsf{causes}.\mathit{BactPericarditis}$
	$\mathit{BactPericarditis} \equiv \mathsf{Inflammation} \sqcap \exists \mathsf{has\_loc}.\mathsf{Pericardium} \sqcap \exists \mathsf{caused\_by}.\mathit{BactInfection}$
$\mathcal{F}$:	$\mathsf{Disease} \sqcap \exists \mathsf{has\_loc}.\mathsf{ConnTissue} \sqsubseteq \exists \mathsf{acts\_on}.\mathsf{ConnTissue}$
	$\mathsf{Inflammation} \sqsubseteq \mathsf{Disease}$
	$\mathsf{Pericardium} \sqsubseteq \mathsf{ConnTissue}$

Fig. 1. Example hybrid $\mathcal{EL}$-TBox

pericarditis'. For instance, bacterial Pericarditis is defined as an inflammation located in the Pericardium caused by a bacterial infection. Note that $\mathcal{T}$ is cyclic. For the primitive concepts in $\mathcal{T}$, the foundation $\mathcal{F}$ states, e.g., that a disease located in connective tissue acts on connective tissue.

The hybrid TBox $(\mathcal{F}, \mathcal{T})$ from Example 1 can be normalized in three steps. Firstly, the first GCI in $\mathcal{F}$ has to be normalized to, e.g.,

$$\begin{aligned} \exists \mathsf{has\_loc}.\mathsf{ConnTissue} &\sqsubseteq \mathsf{HasLocConnTissue} \\ \mathsf{ActsOnConnTissue} &\sqsubseteq \exists \mathsf{acts\_on}.\mathsf{ConnTissue} \\ \mathsf{Disease} \sqcap \mathsf{HasLocConnTissue} &\sqsubseteq \mathsf{ActsOnConnTissue}. \end{aligned}$$

Secondly, $\mathcal{T}$ has to be extended by a definition of the form $A_P \equiv P$ for the primitive concepts Disease, ConnTissue, Infection, Inflammation, Pericardium and also for HasLocConnTissue and ActsOnConnTissue. Thirdly, the primitive names ConnTissue and Pericardium occurring in $\mathcal{T}$ have to be replaced by $A_{\mathsf{ConnTissue}}$ and $A_{\mathsf{Pericardium}}$, respectively.

Normalization serves as an internal preprocessing step to classification and does not replace the original hybrid TBox from the perspective of the user of a DL system. Having introduced hybrid TBoxes formally, it remains to define an appropriate semantics for them.

Definition 17. *Let $(\mathcal{F}, \mathcal{T})$ be a hybrid TBox over N_{prim}, N_{role}, and N_{def}. A primitive interpretation $\mathcal{J}$ is a* model *of $\mathcal{F}$ ($\mathcal{J} \models \mathcal{F}$) iff $C^J \subseteq D^J$ for every GCI $C \sqsubseteq D$ in $\mathcal{F}$. A model $\mathcal{I} \in \text{Int}(\mathcal{J})$ is a* gfp-model *of $(\mathcal{F}, \mathcal{T})$ iff $\mathcal{J} \models \mathcal{F}$ and $\mathcal{I}$ is a gfp-model of $\mathcal{T}$.*

Note that $\mathcal{F}$ ("foundation") is interpreted w.r.t. descriptive semantics while $\mathcal{T}$ ("terminology") is interpreted w.r.t. gfp-semantics. Note also that every gfp-model of $(\mathcal{F}, \mathcal{T})$ can be expressed as $\text{gfp}(\mathcal{T}, \mathcal{J})$ for some primitive interpretation $\mathcal{J}$ with $\mathcal{J} \models \mathcal{F}$.

In order to complete the semantics of hybrid $\mathcal{EL}$-TBoxes, it remains to introduce an appropriate notion of subsumption.

Definition 18. *Let* $(\mathcal{F}, \mathcal{T})$ *be a hybrid* $\mathcal{EL}$*-TBox over* N_{prim}*,* N_{role}*, and* N_{def}*. Let* A, B *be defined concepts in* $\mathcal{T}$*. Then* A is subsumed by B w.r.t. $(\mathcal{F}, \mathcal{T})$ $(A \sqsubseteq_{\text{gfp},\mathcal{F},\mathcal{T}} B)$ *iff* $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ *for all gfp-models* $\mathcal{I}$ *of* $(\mathcal{F}, \mathcal{T})$*.*

For Example 1, we shall see that the subsumption $\mathit{BactPericarditis} \sqsubseteq_{\text{gfp},\mathcal{F},\mathcal{T}}$ $\mathit{ConnTissDisease}$ holds, i.e., Pericarditis is classified as a disease of the connective tissue. How subsumption w.r.t. hybrid TBoxes can be decided in general is the subject of the following section.

Observe that hybrid TBoxes generalize cyclic TBoxes with gfp-semantics in the sense that every cyclic $\mathcal{EL}$-TBox $\mathcal{T}$ can be viewed as a hybrid TBox with an empty foundation. Thus, gfp-subsumption w.r.t. $\mathcal{T}$ coincides with subsumption w.r.t. the hybrid TBox $(\emptyset, \mathcal{T})$. Also note that, every general TBox $\mathcal{T}'$ can be seen as a hybrid TBox $(\mathcal{T}', \emptyset)$. In this case, a descriptive subsumption $P \sqsubseteq_{\mathcal{T}'} Q$ holds iff A_P is subsumed by A_Q w.r.t. the normalized instance of $(\mathcal{T}', \emptyset)$.

3.2 Deciding Subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes

In this section we show that subsumption w.r.t hybrid $\mathcal{EL}$-TBoxes $(\mathcal{F}, \mathcal{T})$ can be reduced to subsumption w.r.t. cyclic $\mathcal{EL}$-TBoxes interpreted by gfp-semantics. The underlying idea is to use the *descriptive* subsumption relations induced by the GCIs in $\mathcal{F}$ to extend the definitions in $\mathcal{T}$ accordingly. To this end, we view the union of $\mathcal{F}$ and $\mathcal{T}$ as a general TBox and ask for all descriptive implications in $\mathcal{T}$ directly involving names from $\mathcal{F}$. These implications are then added to the definitions in $\mathcal{T}$. This notion is formalized as follows.

Definition 19. *Let* $(\mathcal{F}, \mathcal{T})$ *be a normalized hybrid* $\mathcal{EL}$*-TBox over* N_{prim}*,* N_{role}*, and* N_{def}*. For every* $A \in N_{\text{def}}^{\mathcal{T}}$*, let*

$$f(A) := \bigsqcap_{P \in \{P' \in N_{\text{prim}}^{\mathcal{F}} \mid A \sqsubseteq_{\mathcal{F} \cup \mathcal{T}} P'\}} P \sqcap \bigsqcap_{r \in N_{\text{role}}^{\mathcal{F}}} \bigsqcap_{Q \in \{Q' \in N_{\text{prim}}^{\mathcal{F}} \mid A \sqsubseteq_{\mathcal{F} \cup \mathcal{T}} \exists r.Q'\}} \exists r.A_Q \ .$$

The $\mathcal{F}$-completion $f(\mathcal{T})$ *extends the definitions in* $\mathcal{T}$ *as follows.*

$$f(\mathcal{T}) := \{A \equiv C \sqcap f(A) \mid A \equiv C \in \mathcal{T}\}$$

Note that $f(\mathcal{T})$ is still a normalized $\mathcal{EL}$-TBox. To preserve normalization, $f(A)$ adds $\exists r.A_Q$ instead of $\exists r.Q$ whenever A implies $\exists r.Q$.

Example 2. Consider the hybrid TBox $(\mathcal{F}, \mathcal{T})$ from Example 1 after normalization. Our goal is to compute the $\mathcal{F}$-completion of $\mathcal{T}$. To this end, for every defined concept in $\mathcal{T}$, we need to find all descriptive consequences of the form P and $\exists r.P$ implied by $\mathcal{F} \cup \mathcal{T}$, where $P \in N_{\text{prim}}^{\mathcal{F}}$. Obviously, $A_{\mathsf{Inflammation}}$ implies $\mathsf{Disease}$ and $A_{\mathsf{Pericardium}}$ implies $\mathsf{ConnTissue}$. Moreover, $A_{\mathsf{ActsOnConnTissue}}$ yields $\exists \mathsf{acts\_on}.\mathsf{ConnTissue}$. Finally, it is easy to check that *BactPericarditis* implies both $\mathsf{Disease}$ and $\mathsf{HasLocConnTissue}$, and therefore also $\mathsf{ActsOnConnTissue}$, yielding $\exists \mathsf{acts\_on}.\mathsf{ConnTissue}$.

Using these descriptive consequences, the completion $f(A)$ can be computed for every defined name A. Figure 2 shows the "interesting" part of the resulting description graph $\mathcal{G}_{f(\mathcal{T})}$ of the $\mathcal{F}$-completion $f(\mathcal{T})$, omitting some isolated

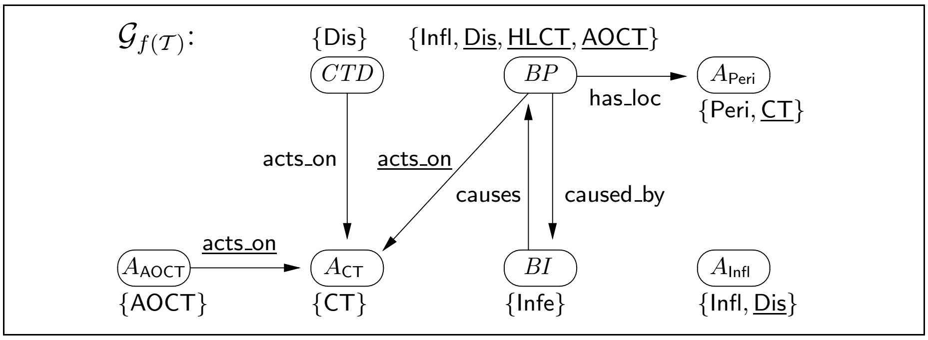

Fig. 2. Example $\mathcal{EL}$-description graph

vertices. Long concept names are abbreviated, i.e., the vertex A_{AOCT} stands for the concept $A_{\mathsf{ActsOnConnTissue}}$, A_{HLCT} for $A_{\mathsf{HasLocConnTissue}}$ and so on. For every vertex A, the label set $L_{f(\mathcal{T})}(A)$ is denoted above or underneath the relevant vertex. Underlined entries are descriptive consequences absent in the original TBox $\mathcal{T}$. As $L_{f(\mathcal{T})}(CTD) \subseteq L_{f(\mathcal{T})}(BP)$ and as BP has the same successor w.r.t. the edge acts_on, it is easy to check that there exists a simulation relation Z on $\mathcal{G}_{f(\mathcal{T})}$ with $(CTD, BP) \in Z$. Therefore, *BactPericarditis* is subsumed by *ConnTissueDisease* w.r.t. $f(\mathcal{T})$ interpreted with gfp-semantics.

Our goal now is to show for a given hybrid $\mathcal{EL}$-TBox $(\mathcal{F}, \mathcal{T})$ and arbitrary names A, B defined in $\mathcal{T}$ that B subsumes A w.r.t. $(\mathcal{F}, \mathcal{T})$ if and only if B subsumes A w.r.t. the $\mathcal{F}$-completion of $\mathcal{T}$ interpreted by gfp-semantics. To this end, we first show that $(\mathcal{F}, \mathcal{T})$ and the $\mathcal{F}$-completed hybrid TBox $(\mathcal{F}, f(\mathcal{T}))$ induce the same subsumption relations.

Lemma 2. *Let $(\mathcal{F}, \mathcal{T})$ be a normalized hybrid $\mathcal{EL}$-TBox over N_{prim}, N_{role}, and N_{def}. Let $A, B \in N_{\text{def}}^{\mathcal{T}}$. Then, $A \sqsubseteq_{\text{gfp},\mathcal{F},\mathcal{T}} B$ iff $A \sqsubseteq_{\text{gfp},\mathcal{F},f(\mathcal{T})} B$.*

Proof. We show that $\text{gfp}(\mathcal{T}, \mathcal{J}) = \text{gfp}(f(\mathcal{T}), \mathcal{J})$, implying for every primitive interpretation $\mathcal{J}$ with $\mathcal{J} \models \mathcal{F}$ that $A^{\text{gfp}(\mathcal{T},\mathcal{J})} \subseteq B^{\text{gfp}(\mathcal{T},\mathcal{J})}$ iff $A^{\text{gfp}(f(\mathcal{T}),\mathcal{J})} \subseteq B^{\text{gfp}(f(\mathcal{T}),\mathcal{J})}$, implying the proposition.

In order to show $\text{gfp}(\mathcal{T}, \mathcal{J}) = \text{gfp}(f(\mathcal{T}), \mathcal{J})$, it suffices to show that, firstly, every model $\mathcal{I} \in \text{Int}(\mathcal{J})$ of $\mathcal{T}$ is also a model of $f(\mathcal{T})$, implying $\text{gfp}(\mathcal{T}_f, \mathcal{J}) \succeq_{\mathcal{J}} \text{gfp}(\mathcal{T}, \mathcal{J})$; and secondly, $\text{gfp}(\mathcal{T}_f, \mathcal{J}) \preceq_{\mathcal{J}} \text{gfp}(\mathcal{T}, \mathcal{J})$.

Consider some model $\mathcal{I} \in \text{Int}(\mathcal{J})$ with $\mathcal{I} \models \mathcal{T}$ and an arbitrary $A \in N_{\text{def}}^{\mathcal{T}}$. As $\mathcal{I} \models \mathcal{T}$, $A^{\mathcal{I}} = \text{deft}_{\mathcal{T}}(A)^{\mathcal{I}}$. Since also $\mathcal{J} \models \mathcal{F}$, $\mathcal{I}$ respects all descriptive implications of $\mathcal{F}$. Hence, we have $A^{\mathcal{I}} \subseteq f(A)^{\mathcal{I}}$, implying $A^{\mathcal{I}} = \text{deft}_{\mathcal{T}}(A)^{\mathcal{I}} \cap f(A)^{\mathcal{I}} = (\text{deft}_{\mathcal{T}}(A) \sqcap f(A))^{\mathcal{I}}$. By definition of $f(\mathcal{T})$, this yields $\mathcal{I} \models f(\mathcal{T})$.

We show $\text{gfp}(f(\mathcal{T}), \mathcal{J}) \preceq_{\mathcal{J}} \text{gfp}(\mathcal{T}, \mathcal{J})$ by transfinite induction on the fixpoint iteration. By Corollary 1, there exists an ordinal α such that $\text{gfp}(f(\mathcal{T}), \mathcal{J}) = \mathcal{I}_{f(\mathcal{T}),\mathcal{J}}^{\downarrow\alpha}$ and $\text{gfp}(\mathcal{T}, \mathcal{J}) = \mathcal{I}_{\mathcal{T},\mathcal{J}}^{\downarrow\alpha}$. We distinguish the case of α being a successor or a limit ordinal.

(α successor ordinal). Induction base: if $\alpha = 0$ then $\mathcal{I}^{\downarrow\alpha}_{f(\mathcal{T}),\mathcal{J}} = \mathcal{I}^{\text{top}} = \mathcal{I}^{\downarrow\alpha}_{\mathcal{T},\mathcal{J}}$, implying $\mathcal{I}^{\downarrow 0}_{f(\mathcal{T}),\mathcal{J}} \preceq_{\mathcal{J}} \mathcal{I}^{\downarrow 0}_{\mathcal{T},\mathcal{J}}$. Induction step: for every $\beta < \alpha$, assume (IH) that $\mathcal{I}^{\downarrow\beta}_{f(\mathcal{T}),\mathcal{J}} \preceq_{\mathcal{J}} \mathcal{I}^{\downarrow\beta}_{\mathcal{T},\mathcal{J}}$. Consider an arbitrary $A \in N^{\mathcal{T}}_{\text{def}}$ defined in $\mathcal{T}$ by

$$A \equiv P_1 \sqcap \cdots \sqcap P_m \sqcap \exists r_1.B_1 \sqcap \ldots \exists r_\ell.B_\ell.$$

We have to show $A^{\mathcal{I}^{\downarrow\beta+1}_{f(\mathcal{T}),\mathcal{J}}} \subseteq A^{\mathcal{I}^{\downarrow\beta+1}_{\mathcal{T},\mathcal{J}}}$. The concept name A is interpreted by $\mathcal{I}^{\downarrow\beta+1}_{f(\mathcal{T}),\mathcal{J}}$ as

$$A^{\mathcal{I}^{\downarrow\beta+1}_{f(\mathcal{T}),\mathcal{J}}} = \bigcap_{1\leq i\leq m} P_i^{\mathcal{J}} \cap \bigcap_{1\leq j\leq \ell} (\exists r_j.B_j)^{\mathcal{I}^{\downarrow\beta}_{f(\mathcal{T}),\mathcal{J}}} \cap f(A)^{\mathcal{I}^{\downarrow\beta}_{f(\mathcal{T}),\mathcal{J}}}.$$

For the original TBox $\mathcal{T}$ we analogously have

$$A^{\mathcal{I}^{\downarrow\beta+1}_{\mathcal{T},\mathcal{J}}} = \bigcap_{1\leq i\leq m} P_i^{\mathcal{J}} \cap \bigcap_{1\leq j\leq \ell} (\exists r_j.B_j)^{\mathcal{I}^{\downarrow\beta}_{\mathcal{T},\mathcal{J}}}.$$

Hence, it suffices to show for every $r \in N^{\mathcal{T}}_{\text{role}}$ and every $B \in N^{\mathcal{T}}_{\text{def}}$ that the subset relation $(\exists r.B)^{\mathcal{I}^{\downarrow\beta}_{f(\mathcal{T}),\mathcal{J}}} \subseteq (\exists r.B)^{\mathcal{I}^{\downarrow\beta}_{\mathcal{T},\mathcal{J}}}$ holds. By (IH), $B^{\mathcal{I}^{\downarrow\beta}_{f(\mathcal{T}),\mathcal{J}}} \subseteq B^{\mathcal{I}^{\downarrow\beta}_{\mathcal{T},\mathcal{J}}}$. As $r^{\mathcal{I}^{\downarrow\beta}_{f(\mathcal{T}),\mathcal{J}}} = r^{\mathcal{J}} = r^{\mathcal{I}^{\beta}_{\mathcal{T},\mathcal{J}}}$, the subset relation immediately carries over to the interpretations of $\exists r.B$.

(α *limit ordinal*). Assume (IH) that $\mathcal{I}^{\downarrow\beta}_{f(\mathcal{T}),\mathcal{J}} \preceq_{\mathcal{J}} \mathcal{I}^{\downarrow\beta}_{\mathcal{T},\mathcal{J}}$ for every $\beta < \alpha$. By definition, in the limit ordinal case it holds for every $B \in N^{\mathcal{T}}_{\text{def}}$ that $B^{\mathcal{I}^{\downarrow\alpha}_{f(\mathcal{T}),\mathcal{J}}}$ equals $\bigcap_{\beta<\alpha} B^{\mathcal{I}^{\downarrow\beta}_{f(\mathcal{T}),\mathcal{J}}}$ which due to (IH) is a subset of $\bigcap_{\beta<\alpha} B^{\mathcal{I}^{\downarrow\beta}_{\mathcal{T},\mathcal{J}}}$ which in turn equals $B^{\mathcal{I}^{\downarrow\alpha}_{\mathcal{T},\mathcal{J}}}$. ∎

Hence, the $\mathcal{F}$-completion $(\mathcal{F}, f(\mathcal{T}))$ of preserves the same subsumption relations as the original. The next lemma shows that, after $\mathcal{F}$-completing $\mathcal{T}$, we may 'forget' $\mathcal{F}$ and still obtain the same subsumptions.

Lemma 3. *Let $(\mathcal{F}, \mathcal{T})$ be a normalized hybrid $\mathcal{EL}$-TBox over N_{prim}, N_{role}, and N_{def}. Let $A, B \in N^{\mathcal{T}}_{\text{def}}$. Then, $A \sqsubseteq_{\text{gfp},\mathcal{F},f(\mathcal{T})} B$ iff $A \sqsubseteq_{\text{gfp},f(\mathcal{T})} B$*

Proof. ($\Leftarrow$) trivial. ($\Rightarrow$) Assume $A \not\sqsubseteq_{\text{gfp},f(\mathcal{T})} B$. We construct a countermodel showing $A \not\sqsubseteq_{\text{gfp},\mathcal{F},f(\mathcal{T})} B$, i.e., a primitive interpretation $\mathcal{J}$ with $\mathcal{J} \models \mathcal{F}$ and $A^{\text{gfp}(f(\mathcal{T}),\mathcal{J})} \not\subseteq B^{\text{gfp}(f(\mathcal{T}),\mathcal{J})}$.

Define $\mathcal{J} =: (\Delta^{\mathcal{J}}, \cdot^{\mathcal{J}})$ as follows.

- $\Delta^{\mathcal{J}} := \{x_A \mid A \in N^{f(\mathcal{T})}_{\text{def}}\}$;
- $P^{\mathcal{J}} := \{x_A \in \Delta^{\mathcal{J}} \mid P \in \text{def}_{f(\mathcal{T})}(A)\}$ for all $P \in N^{f(\mathcal{T})}_{\text{prim}}$;
- $r^{\mathcal{J}} := \{(x_A, x_B) \in (\Delta^{\mathcal{J}})^2 \mid \exists r.B \in \text{def}_{f(\mathcal{T})}(A)\}$ for all $r \in N^{f(\mathcal{T})}_{\text{role}}$.

We first show $x_A \in A^{\text{gfp}(f(\mathcal{T}),\mathcal{J})} \setminus B^{\text{gfp}(f(\mathcal{T}),\mathcal{J})}$. By Proposition 2 it suffices to find a simulation relation $Z\colon \mathcal{G}_{f(\mathcal{T})} \rightleftarrows \mathcal{G}_{\mathcal{J}}$ with $(A, x_a) \in Z$. Define $Z := \{(A, x_A) \mid$

$A \in N_{\mathrm{def}}^{f(\mathcal{T})}\}$. As obviously $(A, x_A) \in Z$, it remains to show that Z respects Definition 14. (S1) For every $(A, x_A) \in Z$, $L_{\mathcal{G}_{f(\mathcal{T})}}(A)$ equals $\mathrm{def}_{f(\mathcal{T})}(A) \cap N_{\mathrm{prim}}^{f(\mathcal{T})}$ which equals $\{P \in N_{\mathrm{prim}}^{f(\mathcal{T})} \mid P \in \mathrm{def}_{f(\mathcal{T})}(A)\} = L_{\mathcal{G}_{\mathcal{J}}}(A)$. (S2) If $(A, x_A) \in Z$ and $(A, r, B) \in E_{\mathcal{G}_{f(\mathcal{T})}}$ then $\exists r.B \in \mathrm{def}_{f(\mathcal{T})}(A)$, implying $(x_A, x_B) \in r^{\mathcal{J}}$, implying $(x_A, r, x_B) \in E_{\mathcal{G}_{\mathcal{J}}}$. Moreover, $(B, x_B) \in Z$. Hence, by (S1) and (S2), $Z\colon \mathcal{G}_{f(\mathcal{T})} \rightleftarrows \mathcal{G}_{\mathcal{J}}$.

Observe that under (S1) we proved *equality* of $L_{\mathcal{G}_{f(\mathcal{T})}}(A)$ and $L_{\mathcal{G}_{\mathcal{J}}}(A)$. Moreover, (S2) also holds in the direction from $\mathcal{G}_{\mathcal{J}}$ to $\mathcal{G}_{f(\mathcal{T})}$: whenever $(A, x_A) \in Z$ and $(x_A, r, x_B) \in E_{\mathcal{G}_{\mathcal{J}}}$ then $(A, r, B) \in E_{\mathcal{G}_{f(\mathcal{T})}}$. Hence, $Z^{-1}\colon \mathcal{G}_{\mathcal{J}} \rightleftarrows \mathcal{G}_{f(\mathcal{T})}$.

Assume $x_A \in B^{\mathrm{gfp}(f(\mathcal{T}),\mathcal{J})}$. Then, by Proposition 2, there is a simulation relation $Y\colon \mathcal{G}_{f(\mathcal{T})} \rightleftarrows \mathcal{G}_{\mathcal{J}}$ with $(B, x_A) \in Y$. But then, by Lemma 1, $Y \circ Z^{-1}$ is a simulation relation on $\mathcal{G}_{f(\mathcal{T})}$ with $(B, A) \in Y \circ Z^{-1}$, implying $A \sqsubseteq_{\mathrm{gfp}, f(\mathcal{T})} B$, in contradiction to the assumption. It remains to show that $J \models \mathcal{F}$. As $\mathcal{F}$ is normalized, we have four types of GCIs in $\mathcal{F}$.

1. $P \sqsubseteq Q \in \mathcal{F}$. If $x_A \in P^{\mathcal{J}}$ then $P \in \mathrm{def}_{f(\mathcal{T})}(A)$, implying $A \sqsubseteq_{\mathcal{F} \cup \mathcal{T}} Q$ which implies $f(A) \sqsubseteq Q$. Hence, $Q \in \mathrm{def}_{f(\mathcal{T})}(A)$, implying $x_A \in Q^{\mathcal{J}}$.
2. $P_1 \sqcap P_2 \sqsubseteq Q \in \mathcal{F}$. If $x_A \in P_1^{\mathcal{J}} \cap P_2^{\mathcal{J}}$ then $P_1, P_2 \in \mathrm{def}_{f(\mathcal{T})}(A)$. This implies $A \sqsubseteq_{\mathcal{F} \cup \mathcal{T}} Q$ which analogously yields $x_A \in Q^{\mathcal{J}}$.
3. $P \sqsubseteq \exists r.Q \in \mathcal{F}$. If $x_A \in P^{\mathcal{J}}$ then $P \in \mathrm{def}_{f(\mathcal{T})}(A)$, implying $A \sqsubseteq_{\mathcal{F} \cup \mathcal{T}} \exists r.Q$. Hence, $\exists r.A_Q \in \mathrm{def}_{f(\mathcal{T})}(A)$, implying $(x_A, r, x_{A_{\exists r.Q}}) \in r^J$. By definition, $Q \in \mathrm{def}_{f(\mathcal{T})}(A_{\exists r.Q})$, implying $x_{A_{\exists r.Q}} \in Q^{\mathcal{J}}$.
4. $\exists r.Q \sqsubseteq P \in \mathcal{F}$. If $x_A \in (\exists r.Q)^{\mathcal{J}}$ then there exists some $x_B \in \Delta^{\mathcal{J}}$ such that $(x_A, r, x_B) \in r^{\mathcal{J}}$ and $x_B \in Q^{\mathcal{J}}$. Hence, $\exists r.B \in \mathrm{def}_{f(\mathcal{T})}(A)$ and $Q \in \mathrm{def}_{f(\mathcal{T})}(B)$. This implies $A \sqsubseteq_{\mathcal{F} \cup \mathcal{T}} P$, implying $P \in \mathrm{def}_{f(\mathcal{T})}(A)$ which yields $x_A \in P^{\mathcal{J}}$. ∎

As an immediate consequence of Lemmas 2 and 3, we obtain the following theorem summarizing our reduction from hybrid $\mathcal{EL}$-TBoxes to cyclic $\mathcal{EL}$-TBoxes.

Theorem 2. *Let $(\mathcal{F}, \mathcal{T})$ be a hybrid $\mathcal{EL}$-TBox over N_{prim}, N_{role}, and N_{def}. Let $A, B \in N_{\mathrm{def}}^{\mathcal{T}}$. Then, $A \sqsubseteq_{\mathrm{gfp}, \mathcal{F}, \mathcal{T}} B$ iff $A \sqsubseteq_{\mathrm{gfp}, f(\mathcal{T})} B$.*

It remains to show that subsumption w.r.t. hybrid TBoxes can be decided in polynomial time.

Corollary 3. *Subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes can be decided in polynomial time.*

Proof. By Corollary 2, gfp-subsumption w.r.t. cyclic $\mathcal{EL}$-TBoxes can be decided in polynomial time. Hence, given $(\mathcal{F}, \mathcal{T})$, it suffices to show for every $A \in N_{\mathrm{def}}^{\mathcal{T}}$ that $f(A)$ is of polynomial size in the size of $(\mathcal{F}, \mathcal{T})$ and can be computed in polynomial time.

By definition, every concept description $f(A)$ contains only conjuncts of the form P and $\exists r.A_P$ with $P \in N_{\mathrm{prim}}^{\mathcal{T}}$ occurring in $\mathcal{F}$. The size of $f(A)$ is therefore linear in the size of $(\mathcal{F}, \mathcal{T})$. It has been shown in [19], that subsumption w.r.t. general $\mathcal{EL}$-TBoxes can be decided in polynomial time, implying that subsumption w.r.t. $\mathcal{F} \cup \mathcal{T}$ is polynomial. ∎

4 Conclusion

Motivated by the goal to make non-standard inference services available to DL systems supporting general TBoxes, the present paper has introduced hybrid $\mathcal{EL}$-TBoxes in which a general $\mathcal{EL}$-TBox $\mathcal{F}$ provides the foundation for a cyclic $\mathcal{EL}$-TBox $\mathcal{T}$ that uses names from $\mathcal{F}$ as primitive concepts. The reduction from Section 3.2 shows that hybrid $\mathcal{EL}$-TBoxes do not extend the expressive power of cyclic $\mathcal{EL}$-TBoxes with gfp-semantics. However, the explicit separation between definitions and implications valid for *all* definitions often leads to smaller and more readable knowledge bases.

The reduction from Section 3.2 also makes non-standard inferences accessible to hybrid TBoxes. It has been shown in [16] that, w.r.t. cyclic $\mathcal{EL}$-TBoxes interpreted by gfp-semantics, the lcs and msc can be computed in polynomial time. A DL system based on hybrid $\mathcal{EL}$-TBoxes could therefore compute the lcs or msc by first (internally) applying the above reduction to the relevant subset of the TBox and then computing the lcs or msc in the way defined in [16].

The technical motivation for choosing $\mathcal{EL}$ as the underlying DL for hybrid TBoxes is that we obtain a polynomial time subsumption problem and can utilize the non-standard inferences known for cyclic $\mathcal{EL}$-TBoxes with gfp-semantics. Our choice, however, is also motivated by applications of $\mathcal{EL}$-TBoxes in the life sciences. For instance, the widely used medical terminology SNOMED [20] corresponds to an $\mathcal{EL}$-Tbox [21]. Similarly, the Gene Ontology [22] can be represented by an $\mathcal{EL}$-TBox with transitive roles, and large parts of the medical knowledge base GALEN [23] can be expressed by a general $\mathcal{EL}$-TBox with transitive roles.

The above applications give rise to the question whether the polynomiality result for subsumption also holds for hybrid TBoxes defined over extensions of $\mathcal{EL}$. An interesting construct to add might be restricted role value maps (RVMs) of the form $r \circ s \sqsubseteq t$ by which, e.g., transitive roles can be defined. Due to positive results for cyclic $\mathcal{EL}$-TBoxes with gfp-semantics [16] and general $\mathcal{EL}$-TBoxes [24], we strongly conjecture that hybrid $\mathcal{EL}$-TBoxes with restricted RVMs can also be classified in polynomial time. For this extension, however, lcs and msc are not yet available. Extending $\mathcal{EL}$ by the bottom concept ($\bot$) would allow to express disjointness constraints of the form $P \sqcap Q \sqsubseteq \bot$ defining P and Q as mutually exclusive concepts.

References

1. Nardi, D., Brachmann, R.: An introduction to description logics. In: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003) 1–40
2. Rector, A., Nowlan, W., Glowinski, A.: Goals for concept representation in the GALEN project. In: Proc. of SCAMC, Washington (1993) 414–418
3. Rector, A.: Medical informatics. In: The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003) 406–426
4. Horrocks, I., Rector, A.L., Goble, C.A.: A description logic based schema for the classification of medical data. In: Proc. of KRDB'96, CEUR-WS (1996)

5. Horrocks, I.R.: Using an expressive description logic: FaCT or fiction? In: Proc. of KR'98. Morgan Kaufmann Publishers (1998) 636–645
6. Haarslev, V., Möller, R.: RACER system description. Lecture Notes in Computer Science **2083** (2001) 701–712
7. Nebel, B.: Terminological cycles: Semantics and computational properties. In: Principles of Semantic Networks: Explorations in the Representation of Knowledge. Morgan Kaufmann Publishers, San Mateo (CA), USA (1991) 331–361
8. Cohen, W.W., Hirsh, H.: Learning the classic description logic: Theoretical and experimental results. In: Proc. of KR'94, Morgan Kaufmann Publishers (1994) 121–133
9. Frazier, M., Pitt, L.: CLASSIC learning. Machine Learning **25** (1996) 151–193
10. Baader, F., Küsters, R.: Computing the least common subsumer and the most specific concept in the presence of cyclic $\mathcal{ALN}$-concept descriptions. In: Proc. of KI-98. Volume 1504 of Lecture Notes in Computer Science, Bremen, Germany, Springer–Verlag (1998) 129–140
11. McGuinness, D.: Explaining Reasoning in Description Logics. Ph.D. dissertation, Department of Computer Science, Rutgers University, New Brunswick, New Jersey (1996)
12. Borgida, A., McGuinness, D.L.: Asking queries about frames. In: Proc. of KR'96 Morgan Kaufmann Publishers, San Francisco, California (1996) 340–349
13. Baader, F., Küsters, R., Borgida, A., McGuinness, D.: Matching in description logics. Journal of Logic and Computation **9** (1999) 411–447
14. Baader, F., Küsters, R., Molitor, R.: Computing least common subsumers in description logics with existential restrictions. In: Proc. of IJCAI-99, Morgan Kaufmann Publishers (1999) 96–101
15. Brandt, S., Turhan, A.-Y.: Using non-standard inferences in description logics—what does it buy me? In: Proc. of KI-2001 Workshop on Applications of Description Logics (KIDLWS'01), Number 44 in CEUR-WS (2001)
16. Baader, F.: Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In Proc. of IJCAI-03, Morgan Kaufmann Publishers (2003) 319–324
17. Baader, F.: Terminological cycles in a description logic with existential restrictions. In: Proc. of IJCAI-03, Morgan Kaufmann Publishers (2003) 325–330
18. Tarski, A.: A lattice-theoretic fixpoint theorem and its applications. Pacific Journal of Mathematics **5** (1955) 285–309
19. Brandt, S.: Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In Proc. of ECAI-2004, IOS Press (2004) 298–302
20. Cote, R., Rothwell, D., Palotay, J., Beckett, R., Brochu, L.: The systematized nomenclature of human and veterinary medicine. Technical report, SNOMED International, Northfield, IL (1993)
21. Spackman, K.: Normal forms for description logic expressions of clinical concepts in SNOMED RT. Journal of the American Medical Informatics Association (2001)
22. Consortium, T.G.O.: Gene Ontology: Tool for the unification of biology. Nature Genetics **25** (2000) 25–29
23. Rector, A., Bechhofer, S., Goble, C.A., Horrocks, I., Nowlan, W.A., Solomon, W.D.: The GRAIL concept modelling language for medical terminology. Artificial Intelligence in Medicine **9** (1997) 139–171
24. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proc. of IJCAI-05, Edinburgh, UK, Morgan Kaufmann Publishers (2005). To appear.