A Proof-Theoretic Approach to Deciding Subsumption and Computing Least Common Subsumer in $\mathcal{EL}$ w.r.t. Hybrid TBoxes

Novak Novaković

Theoretical Computer Science, TU Dresden, Germany

Abstract. Hybrid $\mathcal{EL}$-TBoxes combine general concept inclusions (GCIs), which are interpreted with descriptive semantics, with cyclic concept definitions, which are interpreted with greatest fixpoint (gfp) semantics. We introduce a proof-theoretic approach that yields a polynomial-time decision procedure for subsumption, and present a proof-theoretic computation of least common subsumers in $\mathcal{EL}$ w.r.t. hybrid TBoxes.

1 Introduction

The $\mathcal{EL}$-family of description logics (DLs) is a family of inexpressive DLs whose main distinguishing feature is that they provide their users with existential restrictions rather than value restrictions as the main concept constructor involving roles. The core language of this family is $\mathcal{EL}$, which has the top concept ($\top$), conjunction ($\sqcap$), and existential restrictions ($\exists r.C$) as concept constructors. This family has recently drawn considerable attention since, on the one hand, the subsumption problem stays tractable (i.e., decidable in polynomial time) in situations where the corresponding DL with value restrictions becomes intractable. In particular, subsumption in $\mathcal{EL}$ is tractable both w.r.t. cyclic TBoxes interpreted with gfp or descriptive semantics [3] and w.r.t. general TBoxes (i.e., finite sets of GCIs) interpreted with descriptive semantics [6,4]. On the other hand, although of limited expressive power, $\mathcal{EL}$ is nevertheless used in applications, e.g., to define biomedical ontologies. For example, both the large medical ontology Snomed ct [14] and the Gene Ontology [1] can be expressed in $\mathcal{EL}$, and the same is true for large parts of the medical ontology Galen [12].

In some cases, it would be advantageous to have both GCIs interpreted with descriptive semantics and cyclic concept definitions interpreted with gfp-semantics available in one TBox. One motivation for such hybrid TBoxes comes from the area of non-standard inferences in DLs. For example, if one wants to support the so-called bottom-up construction of DL knowledge bases, then one needs to compute least common subsumers (lcs) and most specific concepts (msc) [5]. In [2], it was shown that the lcs and the msc in $\mathcal{EL}$ always exist and can be computed in polynomial time if cyclic definitions that are interpreted with gfp-semantics are available. In contrast, if cyclic definitions or GCIs are interpreted

S. Hölldobler, C. Lutz, and H. Wansing (Eds.): JELIA 2008, LNAI 5293, pp. 311–323, 2008.

with descriptive semantics, neither the lcs nor the msc need to exist. Hybrid $\mathcal{EL}$-TBoxes have first been introduced in [8]. Basically, such a TBox consists of two parts $\mathcal{T}$ and $\mathcal{F}$, where $\mathcal{T}$ is a cyclic TBox whose primitive concepts occur in the GCIs of the general TBox $\mathcal{F}$. However, defined concepts of $\mathcal{T}$ must not occur in $\mathcal{F}$. It was shown in [8] that subsumption w.r.t. such hybrid TBoxes can still be decided in polynomial time. The algorithm uses reasoning w.r.t. the general TBox $\mathcal{F}$ to extend the cyclic TBox $\mathcal{T}$ to a cyclic TBox $\widehat{\mathcal{T}}$ such that subsumption can then be decided considering only $\widehat{\mathcal{T}}$. In [7] it was shown that, w.r.t. hybrid $\mathcal{EL}$-TBoxes, the lcs and msc always exists and can be computed in polynomial time.

Both, the existing algorithm for deciding subsumption ([8]), and the algorithms for computing lcs and mcs ([7]) in $\mathcal{EL}$ w.r.t. hybrid TBoxes include a pre-processing step of *normalization* of the terminologies. Normalization is, consequently, also required for the algorithms for deciding subsumption, and the algorithms for computing lcs and mcs in $\mathcal{EL}$ w.r.t. descriptive and greatest fixpoint semantics from [13]. This pre-processing step has two undesirable features. From the complexity point of view, it causes quadratic blow-up of the terminologies, and thus, a quadratic blowup in the size of the input to the algorithms. Even more important, especially in the cases of lcs and mcs, normalization replaces the original concept definitions from the terminologies by new ones, by introducing new concept names that occur in those modified definitions. For instance, assume one wants to extend an existing large life-science ontology (and those are usually not normalized) by adding just a single lcs of some two defined concepts. The existing procedure results in quadratic blow-up of the entire ontology, its modification for all users of the ontology, and introduction of some new (generic and unintuitive) concept names.

An approach for deciding subsumption in $\mathcal{EL}$ that significantly differs from the ones described in [3,6,4] was introduced in [9]. It is based on sound and complete Gentzen-style proof calculi for subsumption w.r.t. cyclic TBoxes interpreted with gfp semantics and for subsumption w.r.t. general TBoxes interpreted with descriptive semantics. These calculi yield polynomial-time decision procedures since they satisfy an appropriate sub-description property.

This paper shows that a polynomial-time decision procedure can be obtained for deciding subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes by combining the two calculi introduced in [9]. Another contribution of this paper is a proof-theoretic computation of lcs in $\mathcal{EL}$ w.r.t. hybrid TBoxes. In both cases, the normalization of the ontologies is avoided, together with the undesirable features that come along with it.

2 Hybrid $\mathcal{EL}$-TBoxes

Starting with a set N_{con} of concept names and a set N_{role} of role names, $\mathcal{EL}$-*concept descriptions* are built using the concept constructors top concept ($\top$), conjunction ($\sqcap$), and existential restrictions ($\exists r.C$). The semantics of $\mathcal{EL}$ is defined in the usual way, using the notion of an interpretation $\mathcal{I} = (\mathcal{D}_\mathcal{I}, \cdot^\mathcal{I})$, which consists of a nonempty domain $\mathcal{D}_\mathcal{I}$ and an interpretation function $\cdot^\mathcal{I}$ that assigns

Table 1. Syntax and semantics of $\mathcal{EL}$

Name	Syntax	Semantics
concept name	A	$A^{\mathcal{I}} \subseteq \mathcal{D}_{\mathcal{I}}$
role name	r	$r^{\mathcal{I}} \subseteq \mathcal{D}_{\mathcal{I}} \times \mathcal{D}_{\mathcal{I}}$
top-concept	$\top$	$\top^{\mathcal{I}} = \mathcal{D}_{\mathcal{I}}$
conjunction	$C \sqcap D$	$(C \sqcap D)^{\mathcal{I}} = C^{\mathcal{I}} \cap D^{\mathcal{I}}$
exist. restriction	$\exists r.C$	$(\exists r.C)^{\mathcal{I}} = \{x \mid \exists y : (x,y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$
concept definition	$A \equiv C$	$A^{\mathcal{I}} = C^{\mathcal{I}}$
subsumption	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$

binary relations on $\mathcal{D}_{\mathcal{I}}$ to role names and subsets of $\mathcal{D}_{\mathcal{I}}$ to concept descriptions, as shown in the semantics column of Table 1. A *concept definition* is an expression of the form $A \equiv C$, where A is a concept name and C is a concept description, and a *general concept inclusion* (GCI) is an expression of the form $C \sqsubseteq D$, where C, D are concept descriptions. An interpretation $\mathcal{I}$ is a *model* of a concept definition or GCI if it satisfies the respective condition given in the semantics column of Table 1. This semantics for GCIs and concept definitions is usually called *descriptive semantics*. A *TBox* is a finite set $\mathcal{T}$ of concept definitions that does not contain multiple definitions, i.e., $\{A \equiv C, A \equiv D\} \subseteq \mathcal{T}$ implies $C = D$. Note that TBoxes are *not* required to be *acyclic*, i.e., there may be cyclic dependencies among the concept definitions. A *general TBox* is a finite set of GCIs. The interpretation $\mathcal{I}$ is a *model* of the TBox $\mathcal{T}$ (the general TBox $\mathcal{F}$) iff it is a model of all concept definitions (GCIs) in $\mathcal{T}$ (in $\mathcal{F}$). The name *general* TBox is justified by the fact that concept definitions $A \equiv C$ can of course be expressed by GCIs $A \sqsubseteq C, C \sqsubseteq A$. However, in hybrid TBoxes to be considered, concept definitions will be interpreted by greatest fixpoint semantics rather than by descriptive semantics. We assume in the following that the set of concept names N_{con} is partitioned into the set of *primitive concept names* N_{prim} and the set of *defined concept names* N_{def}. In a hybrid TBox, concept names occurring on the left-hand side of a concept definition are required to come from the set N_{def}, whereas GCIs may not contain concept names from N_{def}.

Definition 1 (Hybrid $\mathcal{EL}$-TBoxes). *A* hybrid $\mathcal{EL}$-TBox *is a pair* $(\mathcal{F}, \mathcal{T})$*, where* $\mathcal{F}$ *is a general* $\mathcal{EL}$*-TBox containing only concept names from* N_{prim}*, and* $\mathcal{T}$ *is an* $\mathcal{EL}$*-TBox such that* $A \equiv C \in \mathcal{T}$ *implies* $A \in N_{def}$*.*

An example of a hybrid $\mathcal{EL}$-Tbox, taken from [8], is given in Fig. 1. It defines the concepts 'disease of the connective tissue,' 'bacterial infection,' and 'bacterial pericarditis' using the cyclic definitions in $\mathcal{T}$. The general TBox $\mathcal{F}$ states some properties that the primitive concepts and roles occurring in $\mathcal{T}$ must satisfy, such as the fact that a disease located on connective tissue also acts on the connective tissue. In general, the idea underlying the definition of hybrid TBoxes is the following: $\mathcal{F}$ can be used to constrain the interpretation of the primitive concepts and roles,

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

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

whereas $\mathcal{T}$ tells us how to interpret the defined concepts occurring in it, once the interpretation of the primitive concepts and roles is fixed.

A *primitive interpretation* $\mathcal{J}$ is defined like an interpretation, with the only difference that it does not provide an interpretation for defined concepts. A primitive interpretation can thus interpret concept descriptions built over N_{prim} and N_{role}, but it cannot interpret concept descriptions containing elements of N_{def}. Given a primitive interpretation $\mathcal{J}$, we say that the (full) interpretation $\mathcal{I}$ is *based on* $\mathcal{J}$ if it has the same domain as $\mathcal{J}$ and its interpretation function coincides with $\mathcal{J}$ on N_{prim} and N_{role}.

Given two interpretations $\mathcal{I}_1$ and $\mathcal{I}_2$ based on the same primitive interpretation $\mathcal{J}$, we define

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

It is easy to see that the relation $\preceq_{\mathcal{J}}$ is a partial order on the set of interpretations based on $\mathcal{J}$. In [3] the following was shown: given an $\mathcal{EL}$-TBox $\mathcal{T}$ and a primitive interpretation $\mathcal{J}$, there exists a unique model $\mathcal{I}$ of $\mathcal{T}$ such that

- $\mathcal{I}$ is based on $\mathcal{J}$;
- $\mathcal{I}' \preceq_{\mathcal{J}} \mathcal{I}$ for all models $\mathcal{I}'$ of $\mathcal{T}$ that are based on $\mathcal{J}$.

We call such a model $\mathcal{I}$ a *gfp-model* of $\mathcal{T}$.

Definition 2 (Semantics of hybrid $\mathcal{EL}$-TBoxes). *An interpretation $\mathcal{I}$ is a* hybrid model *of the hybrid $\mathcal{EL}$-TBox $(\mathcal{F}, \mathcal{T})$, iff $\mathcal{I}$ is a gfp-model of $\mathcal{T}$ and the primitive interpretation $\mathcal{J}$ it is based on is a model of $\mathcal{F}$.*

It is well-known that gfp-semantics coincides with descriptive semantics for acyclic TBoxes. Thus, if $\mathcal{T}$ is actually acyclic, then $\mathcal{I}$ is a hybrid model of $(\mathcal{F}, \mathcal{T})$ according to the semantics introduced in Definition 2, iff it is a model of $\mathcal{T} \cup \mathcal{F}$ w.r.t. descriptive semantics, i.e., iff $\mathcal{I}$ is a model of every GCI in $\mathcal{F}$ and of every concept definition in $\mathcal{T}$.

3 Subsumption w.r.t. Hybrid $\mathcal{EL}$-TBoxes

Based on the semantics for hybrid TBoxes introduced above, we can now define the main inference problem that we want to solve in this paper.

Definition 3 (Subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes). *Let $(\mathcal{F}, \mathcal{T})$ be a hybrid $\mathcal{EL}$-TBox, and A, B defined concepts occurring on the left-hand side of a definition in $\mathcal{T}$. Then A* is subsumed by *B* w.r.t. $(\mathcal{F}, \mathcal{T})$ *(written $A \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}} B$), iff $A^\mathcal{I} \subseteq B^\mathcal{I}$ holds for all hybrid models $\mathcal{I}$ of $(\mathcal{F}, \mathcal{T})$.*

Defining (and computing) subsumption only for concept names A, B defined in $\mathcal{T}$ rather than for arbitrary concept descriptions C, D is not a real restriction since one can always add definitions with the right-hand sides C, D to $\mathcal{T}$.

Assume that the hybrid $\mathcal{EL}$-TBox $(\mathcal{F}, \mathcal{T})$ is given, and that we want to decide whether, for given defined concepts A, B, the subsumption relationship $A \sqsubseteq_{\mathrm{gfp},\mathcal{F},\mathcal{T}} B$ holds or not. Following the ideas in [9], we introduce a sound and complete Gentzen-style calculus for subsumption. The reason why this calculus yields a decision procedure is basically that it has the sub-description property, i.e., application of rules can be restricted to sub-descriptions of concept descriptions occurring in $\mathcal{F}$ or $\mathcal{T}$.

A *sequent for* $(\mathcal{F}, \mathcal{T})$ is of the form $C \sqsubseteq_n D$, where C, D are sub-descriptions of concept descriptions occurring in $\mathcal{F}$ or $\mathcal{T}$, and $n \geq 0$. The rules of the **H**ybrid $\mathcal{EL}$-TBox **C**alculus HC depicted in Fig. 2 can be used to derive new sequents from sequents that have already been derived. For example, the sequents in the first row of the figure can always be derived without any prerequisites, using the rules Refl, Top, and Start, respectively. Using the rule AndR, the sequent $C \sqsubseteq_n D \sqcap E$ can be derived in case both $C \sqsubseteq_n D$ and $C \sqsubseteq_n E$ have already been derived. Note that the rule Start applies only for $n = 0$. Also note that, in the rule DefR, the index is incremented when going from the prerequisite to the consequent.

$$
\begin{array}{llllll}
C \sqsubseteq_n C & \text{(Refl)} & C \sqsubseteq_n \top & \text{(Top)} & C \sqsubseteq_0 D & \text{(Start)} \\[2ex]
\dfrac{C \sqsubseteq_n E}{C \sqcap D \sqsubseteq_n E} & \text{(AndL1)} & \dfrac{D \sqsubseteq_n E}{C \sqcap D \sqsubseteq_n E} & \text{(AndL2)} & \dfrac{C \sqsubseteq_n D \quad C \sqsubseteq_n E}{C \sqsubseteq_n D \sqcap E} & \text{(AndR)} \\[2ex]
 & & \dfrac{C \sqsubseteq_n D}{\exists r.C \sqsubseteq_n \exists r.D} & \text{(Ex)} & & \\[2ex]
\dfrac{C \sqsubseteq_n D}{A \sqsubseteq_n D} & \text{(DefL)} & \dfrac{D \sqsubseteq_n C}{D \sqsubseteq_{n+1} A} & \text{(DefR)} & \dfrac{C \sqsubseteq_n E \quad F \sqsubseteq_n D}{C \sqsubseteq_n D} & \text{(GCI)} \\
\text{for } A \equiv C \in \mathcal{T} & & \text{for } A \equiv C \in \mathcal{T} & & \text{for } E \sqsubseteq F \in \mathcal{F} &
\end{array}
$$

Fig. 2. The rule system HC

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\mathsf{Infl} \sqsubseteq_n \mathsf{Infl} \quad \mathsf{D} \sqsubseteq_n \mathsf{D}}{\mathsf{Infl} \sqsubseteq_n \mathsf{D}}}{\mathsf{Infl} \sqcap \exists \mathsf{hl.P} \sqsubseteq_n \mathsf{D}} \quad \dfrac{\dfrac{\dfrac{\dfrac{\mathsf{Infl} \sqsubseteq_n \mathsf{Infl} \quad \mathsf{D} \sqsubseteq_n \mathsf{D}}{\mathsf{Infl} \sqsubseteq_n \mathsf{D}}}{\mathsf{Infl} \sqcap \exists \mathsf{hl.P} \sqsubseteq_n \mathsf{D}} \quad \dfrac{\dfrac{\dfrac{\mathsf{P} \sqsubseteq_n \mathsf{P} \quad \mathsf{CT} \sqsubseteq_n \mathsf{CT}}{\mathsf{P} \sqsubseteq_n \mathsf{CT}}}{\exists \mathsf{hl.P} \sqsubseteq_n \exists \mathsf{hl.CT}}}{\mathsf{Infl} \sqcap \exists \mathsf{hl.P} \sqsubseteq_n \exists \mathsf{hl.CT}}}{\mathsf{Infl} \sqcap \exists \mathsf{hl.P} \sqsubseteq_n \mathsf{D} \sqcap \exists \mathsf{hl.CT}} \quad \exists \mathsf{ao.CT} \sqsubseteq_n \exists \mathsf{ao.CT}}{\mathsf{Infl} \sqcap \exists \mathsf{hl.P} \sqsubseteq_n \exists \mathsf{ao.CT}}}{\mathsf{Infl} \sqcap \exists \mathsf{hl.P} \sqsubseteq_n \mathsf{D} \sqcap \exists \mathsf{ao.CT}}}{\mathsf{Infl} \sqcap \exists \mathsf{hl.P} \sqcap \exists \mathsf{cb.Bl} \sqsubseteq_n \mathsf{D} \sqcap \exists \mathsf{ao.CT}}}{\mathsf{BP} \sqsubseteq_n \mathsf{D} \sqcap \exists \mathsf{ao.CT}}}{\mathsf{BP} \sqsubseteq_{n+1} \mathsf{CTD}}
$$

Fig. 3. An example of a derivation in HC

Fig. 3 shows a derivation in HC w.r.t. the hybrid $\mathcal{EL}$-TBox from Fig. 1, where obvious abbreviations of concept and role names have been made. This derivation tree demonstrates that the sequent $\mathsf{BactPericarditis} \sqsubseteq_{n+1} \mathsf{ConnTissDisease}$ can be derived for every $n \geq 0$. Note that we can also derive $\mathsf{BactPericarditis} \sqsubseteq_0 \mathsf{ConnTissDisease}$ using the rule Start.

The calculus HC defines binary relations $\sqsubseteq_n$ for $n \in \{0, 1, \ldots\} \cup \{\infty\}$ on the set of sub-descriptions of concept descriptions occurring in $\mathcal{F}$ or $\mathcal{T}$:

Definition 4. *Let C, D be sub-descriptions of the concept descriptions occurring in $\mathcal{F}$ or $\mathcal{T}$. Then $C \sqsubseteq_n D$ holds, iff the sequent $C \sqsubseteq_n D$ can be derived using the rules of HC. In addition, $C \sqsubseteq_\infty D$ holds, iff $C \sqsubseteq_n D$ holds for all $n \geq 0$.*

The calculus HC is sound and complete for subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes in the following sense.

Theorem 1 (Soundness and Completeness of HC). *Let $(\mathcal{F}, \mathcal{T})$ be a hybrid $\mathcal{EL}$-TBox, and A, B defined concepts occurring on the left-hand side of a definition in $\mathcal{T}$. Then $A \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}} B$, iff $A \sqsubseteq_\infty B$ holds.*

A detailed proof of this theorem is given in [11]. Though the rules of HC are taken from the sound and complete subsumption calculi introduced in [9] for subsumption w.r.t. cyclic $\mathcal{EL}$-TBoxes interpreted with gfp-semantics and for subsumption w.r.t. general $\mathcal{EL}$-TBoxes interpreted with descriptive semantics, respectively, the proof that their combination is sound and complete for the case of hybrid $\mathcal{EL}$-TBoxes requires non-trivial modifications of the proofs given in [9]. Nevertheless, these proofs appear to be simpler and easier to comprehend than the ones given in [8,10] for the correctness of the reduction-based subsumption algorithm for hybrid $\mathcal{EL}$-TBoxes introduced there.

In our example, we have $\mathsf{BactPericarditis} \sqsubseteq_\infty \mathsf{ConnTissDisease}$, and thus soundness of HC implies that the subsumption relationship $\mathsf{BactPericarditis} \sqsubseteq_{\mathrm{gfp},\mathcal{F},\mathcal{T}} \mathsf{ConnTissDisease}$ holds.

It is not hard to show that $\sqsubseteq_0$ is the universal relation on sub-descriptions of the concept descriptions occurring in $\mathcal{F}$ or $\mathcal{T}$, and that $\sqsubseteq_{n+1} \subseteq \sqsubseteq_n$ holds for all $n \geq 0$ (see [11] for a proof). Thus, to compute $\sqsubseteq_\infty$ we can start with the universal relation $\sqsubseteq_0$, and then compute $\sqsubseteq_1, \sqsubseteq_2, \ldots$, until for some m we have $\sqsubseteq_m = \sqsubseteq_{m+1}$, and thus $\sqsubseteq_m = \sqsubseteq_\infty$. Since the set of sub-descriptions is finite, the computation of each relation $\sqsubseteq_n$ can be done in finite time, and we can be sure that there always exists an m such that $\sqsubseteq_m = \sqsubseteq_{m+1}$. This shows that the calculus HC indeed yields a subsumption algorithm. Even more so, the decision procedure will terminate in polynomial time. This can easily be seen by noticing that we can compute each of the relations $\sqsubseteq_m$ in polynomial time by performing the proof search for each pair of subconcepts of the TBox with caching the intermediate derived subsumption pairs. Notice that the number of subconcepts of the TBox is linear in the size of the TBox. Also, the maximal number m_0 of different $\sqsubseteq_m$ relations corresponds to the case where $\sqsubseteq_i$ and $\sqsubseteq_{i+1}$ differ in a single subsumption pair, for all $i \leq m_0$. Thus, the total number of different $\sqsubseteq_m$ relations is bounded by the number of subsumption pairs +1, i.e. bounded by a quadratic function in the size of the TBox.

A detailed description of the implementation of this decision procedure can be found in [15].

4 Computing Least Common Subsumer in $\mathcal{EL}$ w.r.t. Hybrid TBoxes

This section is dedicated to employing the developed proof-theoretic techniques in calculating and showing the correctness of the computation of the least-common subsumer of two defined concepts with respect to hybrid TBoxes. We start by introducing the notion of a conservative extension of a hybrid $\mathcal{EL}$-TBox.

Definition 5. *Given a hybrid $\mathcal{EL}$-TBox $(\mathcal{F}, \mathcal{T}')$ we say that the hybrid TBox $(\mathcal{F}, \mathcal{T}'')$ is a* conservative extension *of $(\mathcal{F}, \mathcal{T}')$, iff $\mathcal{T}' \subseteq \mathcal{T}''$, and $\mathcal{T}'$ and $\mathcal{T}''$ have the same primitive concepts and roles.*

It is well known that the conservative extensions do not change the set of subsumption pairs (see [2]), i.e. $(\mathcal{F}, \mathcal{T}) \models C \sqsubseteq D$, iff $(\mathcal{F}, \mathcal{T}') \models C \sqsubseteq D$, for all subconcepts C, D occurring in $(\mathcal{F}, \mathcal{T})$. This can also be shown in a proof-theoretic way by noticing that $C \sqsubseteq_\infty D$ can be derived in the HC calculus for $(\mathcal{F}, \mathcal{T})$, iff it can be derived in HC for $(\mathcal{F}, \mathcal{T}')$.

Notice that, for instance, $\sqsubseteq_n$, $\sqsubseteq_\infty$ and N_{def} are defined w.r.t. a concrete TBox. That is why, in the cases where multiple TBoxes are concerned, superscripts are used to specify the appropriate TBoxes. The following definition introduces the notion of least-common subsumer in the hybrid setting.

Definition 6. *(Hybrid lcs) Let $(\mathcal{F}, \mathcal{T}_1)$ be a hybrid $\mathcal{EL}$-TBox and $A, B \in N_{def}^{\mathcal{T}_1}$. Let $(\mathcal{F}, \mathcal{T}_2)$ be a conservative extension of $(\mathcal{F}, \mathcal{T}_1)$ with $Z \in N_{def}^{\mathcal{T}_2}$. Then Z in $(\mathcal{F}, \mathcal{T}_2)$ is a hybrid least-common subsumer (lcs) of A, B in $(\mathcal{F}, \mathcal{T}_1)$, iff the following conditions hold:*

1. *$A \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_2} Z$ and $B \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_2} Z$; and*
2. *if $(\mathcal{F}, \mathcal{T}_3)$ is a conservative extension of $(\mathcal{F}, \mathcal{T}_2)$ and $D \in N_{def}^{\mathcal{T}_3}$ such that $A \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_3} D$ and $B \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_3} D$ then $Z \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_3} D$.*

Concept D from the previous definition is an arbitrary concept defined in some conservative extension $(\mathcal{F}, \mathcal{T}_3)$. It would suffice, though, to restrict D to be arbitrary concept defined in $\mathcal{T}_3 \setminus \mathcal{T}_1$, i.e. it is sufficient to consider only newly defined concepts for testing the condition 2 of the definition above. Indeed, if we want to test whether $Z \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_3} D$ for $D \in N_{def}^{\mathcal{T}_1}$, we can equivalently check whether $Z \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_4} A_D$, where $\mathcal{T}_4$ consists of $\mathcal{T}_2$ with a definition $D \equiv A_D$ for a new concept A_D.

One can observe that, as mentioned in the introduction, the existing algorithm for computing lcs for $\mathcal{EL}$ w.r.t. TBoxes interpreted by greatest fixpoint semantics from [13], and the algorithm for computing lcs for $\mathcal{EL}$ w.r.t. hybrid TBoxes from [7], strictly speaking, do not result in a conservative extension of the original TBox due to the normalization step.

Assume now that, given a hybrid TBox $(\mathcal{F}, \mathcal{T})$, one wants to know the least common subsumer of two defined concepts A and B occurring in a hybrid TBox. We give a definition of an extension of the hybrid TBox which contains definitions of lcs of defined concepts occurring in the original TBox.

Before doing so, consider the set *subcon* of all subconcepts of concept descriptions occurring in the TBox $(\mathcal{F}, \mathcal{T})$ and consider the sets

$$\begin{aligned} ExRest &= \{C \mid C \in \mathit{subcon} \text{ and there is an } r \in N_{role} \text{ such that } \exists r.C \in \mathit{subcon}\}, \\ N_{pair} &= \{(C, D) \mid C, D \in (N_{def}^{\mathcal{T}} \cup ExRest)\}, \text{ and} \\ Prims &= \{C \mid C \in \mathit{subcon} \text{ and } C \text{ does not have elments of } N_{def} \text{ as subconcepts}\}. \end{aligned}$$

Notice that elements of $Prims$ are concept descriptions built using only primitive concept names and role names. Now we define the conservative extension of the TBox as follows.

Definition 7. *Let $(\mathcal{F}, \mathcal{T})$ be a hybrid $\mathcal{EL}$-TBox. A conservative extension $(\mathcal{F}, \mathcal{T}_{lcs})$ of $(\mathcal{F}, \mathcal{T})$ is obtained by adding to the $(\mathcal{F}, \mathcal{T})$ definitions*

$$(C, D) \equiv \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_l \sqcap \exists r_1.(C_1, D_1) \sqcap ... \sqcap \exists r_m.(C_m, D_m)$$

for each $(C, D) \in N_{pair}$, where:

1. *$\theta \in \{\theta_1, ..., \theta_k\}$, iff $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T})} \theta$, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T})} \theta$, and $\theta \in Prims$*
2. *$X \in \{X_1, ..., X_l\}$, iff $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T})} X$, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T})} X$, and $X \in N_{def}^{\mathcal{T}}$;*
3. *$\exists r.(\tau, \sigma) \in \{\exists r_1.(C_1, D_1), ..., \exists r_m.(C_m, D_m)\}$, iff $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T})} \exists r.\tau$, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T})} \exists r.\sigma$ and $(\tau, \sigma) \in N_{pair}$.*

Least common subsumer of two defined concepts A and B occurring in the TBox will be newly defined concept (A, B). Once the $\sqsubseteq_\infty^{(\mathcal{F},\mathcal{T})}$ relation is computed, computation of the extension $(\mathcal{F}, \mathcal{T}_{lcs})$ can be done by a simple computation in polynomial time, and the resulting TBox is indeed a conservative extension of the original one.

In what follows we show that the definition above provides computation of lcs of two given concepts, i.e. that the least common subsumer of two defined concepts A and B occurring in the TBox is the newly defined concept (A, B).

What needs to be shown is that conditions 1 and 2 from Definition 6 hold for all (A, B). In order to do so, we will simplify the discussion by restricting our attention to those conservative extensions from condition 2 for which newly added definitions are of a certain regular structure. Such an assumption, of course, is done without loss of generality.

Definition 8. *We say that a conservative extension* $(\mathcal{F}, \mathcal{T}')$ *of the hybrid* $\mathcal{EL}$*-TBox* $(\mathcal{F}, \mathcal{T})$ *is obtained by adding* normalized definitions modulo $(\mathcal{F}, \mathcal{T})$ *if every definition in* $\mathcal{T}' \setminus \mathcal{T}$ *is of the form:*

$$Z \equiv P_1 \sqcap ... \sqcap P_m \sqcap A_1 \sqcap ... \sqcap A_k \sqcap \exists r_1.B_1 \sqcap ... \sqcap \exists r_n.B_n$$

where P_i *is a primitive concept for every* $i = 1, ..., m$, A_i *is a concept defined in* $\mathcal{T}$, *for every* $i = 1, ..., k$, *and* B_j *is a concept defined in* $\mathcal{T}' \setminus \mathcal{T}$, *for every* $j = 1, ..., n$.

The proof of the following proposition can be found in [11].

Proposition 1. *Let* $(\mathcal{F}, \mathcal{T})$ *be a hybrid* $\mathcal{EL}$*-TBox, and* $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$ *some conservative extension of* $(\mathcal{F}, \mathcal{T})$. *Then, there is a conservative extension* $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$ *of* $(\mathcal{F}, \mathcal{T})$ *obtained by adding normalized definitions modulo* $(\mathcal{F}, \mathcal{T})$ *to it, such that the set of defined concepts in* $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$ *is a subset of the set of defined concepts in* $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$, *and* $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1) \models C \sqsubseteq D$, *iff* $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2) \models C \sqsubseteq D$ *for every two concepts* C *and* D *defined in* $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$.

This proposition shows that one can restrict the attention to the conservative extensions obtained by adding the normalized definitions modulo a TBox when checking for property 2 from the definition of lcs. Indeed, let Φ be a concept defined in a conservative extension $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$ of hybrid TBox $(\mathcal{F}, \mathcal{T})$, such that Φ subsumes both A and B. By the previous proposition, there is a conservative extension $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$ of the TBox $(\mathcal{F}, \mathcal{T})$ by normalized definitions modulo $(\mathcal{F}, \mathcal{T})$, such that (A, B) will be subsumed by Φ w.r.t. $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$, iff (A, B) is subsumed by Φ w.r.t. $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$. In particular, (A, B) will be subsumed by every concept Φ that subsumes both A and B w.r.t. an arbitrary conservative extension of the TBox, iff it is subsumed by every concept Φ that subsume both A and B w.r.t. conservative extensions of the TBox obtained by adding normalized definitions modulo $(\mathcal{F}, \mathcal{T})$.

We continue by introducing a relation $\Vdash$. We say that $C \Vdash D$, iff $C \sqsubseteq_n D$ can be derived for some n using only the rules that consider the left-hand side of a sequent, i.e. (Ref), (AndL1), (AndL2), (DefL) and (GCI). Notice that if $C \sqsubseteq_n D$ can be derived using only those rules for some n, then it can be derived for any n. The following, rather technical lemmas are given here without proofs, which can be found in [11].

Lemma 1. *Suppose that* $n > 0$.

– $F \sqsubseteq_{n+1} A$, *iff* $F \sqsubseteq_n C_A$, *where* $A \equiv C_A$ *is an axiom of the TBox.*
– $F \sqsubseteq_n \exists r.D$, *iff there exist* α, β, ρ *such that* $F \sqsubseteq_n \alpha, \beta \Vdash \exists r.\rho$ *and* $\rho \sqsubseteq_n D$ *for some subconcept* ρ *of the TBox, and* α *and* β *being such that either* $\alpha = \beta = F$ *or* $\alpha \sqsubseteq \beta$ *being a GCI from* $\mathcal{F}$.

Lemma 2. *Let D, C and F be arbitrary subconcepts occurring in a TBox $(\mathcal{F}, \mathcal{T})$. If $D \sqsubseteq_\infty F$ and $F \sqsubseteq_n C$, then $D \sqsubseteq_n C$.*

We are equipped now to show that our newly defined concepts are indeed the subsumers.

Lemma 3. *Let D and C be arbitrary concepts from $N_{def}^{\mathcal{T}} \cup ExRest$. Then, $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} (D, C)$ and $C \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} (D, C)$ for every n.*

Proof. We give a proof of $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} (D, C)$, proof of $C \sqsubseteq_n (D, C)$ is analogous. Proof is carried out by induction on n. For $n = 0$, $D \sqsubseteq_0^{(\mathcal{F},\mathcal{T}_{lcs})} (D, C)$ follows from the rule (Start). Assume now that $D \sqsubseteq_l^{(\mathcal{F},\mathcal{T}_{lcs})} (D, C)$ holds for all $l \leq n$. We prove that $D \sqsubseteq_{n+1}^{(\mathcal{F},\mathcal{T}_{lcs})} (D, C)$. Let

$$(D, C) \equiv \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_u \sqcap \exists s_1.(D_{l_1}, C_{m_1}) \sqcap ... \sqcap \exists s_t.(D_{l_t}, C_{m_t})$$

be the definition of (D, C) in the extended hybrid TBox $(\mathcal{F}, \mathcal{T}_{lcs})$. Lemma 1 applied to this definition yields $D \sqsubseteq_{n+1}^{(\mathcal{F},\mathcal{T}_{lcs})} (D, C)$, iff $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_u \sqcap \exists s_1.(D_{l_1}, C_{m_1}) \sqcap ... \sqcap \exists s_t.(D_{l_t}, C_{m_t})$. Therefore, it is sufficient to show $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_u \sqcap \exists s_1.(D_{l_1}, C_{m_1}) \sqcap ... \sqcap \exists s_t.(D_{l_t}, C_{m_t})$. Due to (AndR), one way to show this is to give a proof of $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \theta_i$ for $i = 1, ..., k$, $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} X_i$ for $i = 1, ..., u$, and $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \exists s_j.(D_{l_j}, C_{m_j})$ for $j = 1, ..., t$.

- $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \theta_i$: by Definition 7, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_{lcs})} \theta_i$. Therefore, by definition of $\sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_{lcs})}$, $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \theta_i$, for $i = 1, ..., k$. Similarly, $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} X_i$.
- $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \exists s_j.(D_{l_j}, C_{m_j})$: by Definition 7, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_{lcs})} \exists s_j.D_{l_j}$, therefore $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \exists s_j.D_{l_j}$. Since both D_{l_j} and C_{m_j} belong to the $N_{def} \cup ExRest$, induction hypothesis can be applied and it yields $D_{l_j} \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} (D_{l_j}, C_{m_j})$. Then, $\exists s_j.D_{l_j} \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \exists s_j.(D_{l_j}, C_{m_j})$ follows by applying the rule (Ex). Since $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_{lcs})} \exists s_j.D_{l_j}$, Lemma 2 yields $D \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \exists s_j.(D_{l_j}, C_{m_j})$.

By definition of $\sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_{lcs})}$, and due to the soundness of HC, both A and B are subsumed by (A, B), for all defined concepts A and B in $(\mathcal{F}, \mathcal{T})$.

We give here another technical property of conservative extensions and $\Vdash^{(\mathcal{F},\mathcal{T})}$ relation.

Lemma 4. *Let $(\mathcal{F}, \mathcal{T}_2)$ be an arbitrary conservative extension of $(\mathcal{F}, \mathcal{T})$. If σ is a subconcept occurring in $(\mathcal{F}, \mathcal{T})$ and $\sigma \Vdash^{(\mathcal{F},\mathcal{T}_2)} \exists r.\tau$, then $\exists r.\tau$ is a subconcept occurring in $(\mathcal{F}, \mathcal{T})$.*

Finally, we show the minimality condition.

Lemma 5. *Let $(\mathcal{F}, \mathcal{T}_2)$ be a conservative extension of $(\mathcal{F}, \mathcal{T}_{lcs})$ by normalized definitions modulo $(\mathcal{F}, \mathcal{T}_{lcs})$. Let D and C be two concepts from $N_{def}^{\mathcal{T}} \cup ExRest$, and let Φ be a concept defined in $\mathcal{T}_2 \setminus \mathcal{T}$. If $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$ and $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$, then $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \Phi$, for every n.*

Proof. Assume

$$\Phi \equiv \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_u \sqcap \exists r_1.\Phi_1 \sqcap ... \sqcap \exists r_l.\Phi_l$$

is a definition in $\mathcal{T}_2 \setminus \mathcal{T}$. Here, θ_i, for $i = 1,..,k$, is an element of $Prims$ (from the definition of $(\mathcal{F}, \mathcal{T}_{lcs})$), (and in the case Φ is defined in $\mathcal{T}_2 \setminus \mathcal{T}_{lcs}$, it is a primitive concept). X_i, for $i = 1,..,u$, is a concept defined in $(\mathcal{F}, \mathcal{T}_{lcs})$ (in $\mathcal{T}$ in the case Φ is defined in $\mathcal{T}_{lcs}$), while Φ_i, for $i = 1,...,l$ is a concept defined in $\mathcal{T}_2 \setminus \mathcal{T}$ (in $\mathcal{T}_{lcs} \setminus \mathcal{T}$ in the case Φ is defined in $\mathcal{T}_{lcs}$). Again, proof is carried out by induction on n. For $n = 0$, $(D, C) \sqsubseteq_0^{(\mathcal{F},\mathcal{T}_2)} \Phi$ follows from the rule (Start).

Assume now that $(D, C) \sqsubseteq_k^{(\mathcal{F},\mathcal{T}_2)} \Phi$ for all $k \leq n$. We prove that $(D, C) \sqsubseteq_{n+1}^{(\mathcal{F},\mathcal{T}_2)} \Phi$. One of the properties of the $\sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})}$ relation, shown in Lemma 1 in our case yields $(D, C) \sqsubseteq_{n+1}^{(\mathcal{F},\mathcal{T}_2)} \Phi$ iff $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_u \sqcap \exists r_1.\Phi_1 \sqcap ... \sqcap \exists r_l.\Phi_l$. Therefore, it suffices to prove $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_u \sqcap \exists r_1.\Phi_1 \sqcap ... \sqcap \exists r_l.\Phi_l$. Again due to (AndR), one way to show this is to give a proof of $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \theta_i$ for $i = 1,...,k$, $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} X_i$ for $i = 1,...,u$, and $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\Phi_j$ for $j = 1,...,l$.

- $(D, C) \sqsubseteq_n \theta_i$: by soundness and completeness of HC, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$ implies $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \theta_i$, similarly, $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \theta_i$, and therefore θ_i occurs on the right-hand side of the definition of (D, C) by Definition 7, since θ_i belongs to $Prims$. Therefore, $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \theta_i$ follows from completeness of the HC calculus and the fact that $(D, C) \sqsubseteq \theta_i$ holds in all models of $(\mathcal{F}, \mathcal{T})$.
- $(D, C) \sqsubseteq_n X_i$: we distinguish two cases
  1. X_i is defined in $\mathcal{T}$: by soundness and completeness of HC, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$ implies $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} X_i$, similarly, $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} X_i$, and therefore X_i occurs on the right-hand side of the definition of (D, C) by Definition 7. Thus, $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} X_i$ follows from completeness of the HC calculus and the fact that $(D, C) \sqsubseteq X_i$ holds in all models of $(\mathcal{F}, \mathcal{T})$.
  2. X_i is defined in $\mathcal{T}_{lcs} \setminus \mathcal{T}$: then, X_i is of the form (γ, δ). By soundness and completeness of HC, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$ implies $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} (\gamma, \delta)$, similarly, $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} (\gamma, \delta)$. Now, the induction hypothesis can be applied, since (γ, δ) is defined in $\mathcal{T}_{lcs} \setminus \mathcal{T} \subseteq \mathcal{T}_2 \setminus \mathcal{T}$, and it yields $(D, C) \sqsubseteq_n (\gamma, \delta)$.
- $(D, C) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\Phi_j$: again, by soundness and completeness of HC, $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$ implies $D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\Phi_j$ and $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$ implies $C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\Phi_j$. By Lemma 1, this means that there exist concepts α, β and ρ and such that

$$D \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \alpha, \quad \beta \Vdash^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\rho, \quad \rho \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi_j$$
$$C \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \alpha_1, \; \beta_1 \Vdash^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\rho_1, \; \rho_1 \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi_j$$

  where $\alpha \sqsubseteq \beta \in \mathcal{F}$ or $D = \alpha = \beta$; $\alpha_1 \sqsubseteq \beta_1 \in \mathcal{F}$ or $C = \alpha_1 = \beta_1$; and ρ and ρ_1 are some concepts occurring in $(\mathcal{F}, \mathcal{T}_2)$.

This further implies $D \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\rho$ by applying rule (Concept) to $D \sqsubseteq_{n}^{(\mathcal{F},\mathcal{T}_2)} \alpha$ and $\beta \sqsubseteq_{n}^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\rho$ for every n. Analogously, $C \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\rho_1$.

Lemma 4 applied to $\beta \Vdash^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\rho$ and $\beta_1 \Vdash^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\rho_1$ yields the fact that $\exists r_j.\rho$ and $\exists r_j.\rho_1$ are concepts from $(\mathcal{F},\mathcal{T})$. Even more, they are from $ExRest$.

Now, the induction hypothesis can be applied to $\rho \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T}_2)} \Phi_j$ and $\rho_1 \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T}_2)} \Phi_j$ to obtain $(\rho,\rho_1) \sqsubseteq_{n}^{(\mathcal{F},\mathcal{T}_2)} \Phi_j$. On the other hand, by Definition 7, $\exists r_j.(\rho,\rho_1)$ is one of the conjuncts in the definition of (D,C). Now, $(D,C) \sqsubseteq_{n}^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\Phi_j$ can be derived from $(\rho,\rho_1) \sqsubseteq_{n}^{(\mathcal{F},\mathcal{T}_2)} \Phi_j$, by applying (Ex) rule, (AndL1) or (AndL2) rules several times and (DefL) in the end.

Again, due to the soundness of derivations in HC, considering defined concepts A, B and the corresponding (A,B), we have that (A,B) is subsumed by every concept defined in $\mathcal{T}_2 \setminus \mathcal{T}$ that subsumes both A and B. (We use notation from the previous lemma.) By the comment after Definition 6, this conclusion is sufficient to show property 2 from the definition of hybrid lcs.

Notice also, that, as shown before, the assumption made on the added definitions within the conservative extensions, namely the assumption of them being normalized modulo the TBox, does not cause loss of generality.

Combined with the previously shown property 1 from the definition of hybrid lcs, this proves the following theorem.

Theorem 2. *The concept description* (A,B) *from the extended hybrid TBox* $(\mathcal{F},\mathcal{T}_{lcs})$ *is a least common subsumer of* A *and* B *w.r.t. the hybrid TBox* $(\mathcal{F},\mathcal{T})$.

5 Conclusion

In this paper, we have described a Gentzen-style calculus for subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes, which is an extension to the case of hybrid TBoxes of the calculi for general TBoxes and for cyclic TBoxes with gfp-semantics that have been introduced in [9]. Based on this calculus, we have developed a polynomial-time decision procedure for subsumption w.r.t. hybrid $\mathcal{EL}$-TBoxes. The second result described in this paper was the proof-theoretic computation of least common subsumers w.r.t. hybrid $\mathcal{EL}$-TBoxes. We provide a technique that avoids the undesirable features of normalization. Since the main motivation for considering hybrid TBoxes was that, w.r.t. them, the lcs and msc always exist, the natural next step is to develop a proof-theoretic approach to computing the msc, and we currently investigate that possibility. Other future work in this direction is to try to extend the described techniques to more expressive DLs from the $\mathcal{EL}$ family.

Acknowledgements. The author would like to thank prof. Franz Baader for introducing me into the topic, his valuable advices and supervision of the work, in general.

References

1. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.: Gene ontology: tool for the unification of biology. Nat Genet. 25(1), 25–29 (2000)
2. Baader, F.: Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In: Proc. IJCAI 2003. Morgan Kaufmann, San Francisco (2003)
3. Baader, F.: Terminological cycles in a description logic with existential restrictions. In: Proc. IJCAI 2003. Morgan Kaufmann, San Francisco (2003)
4. Baader, F., Brandt, S., Lutz, C.: Pushing the $\mathcal{EL}$ envelope. In: Proc. IJCAI 2005. Morgan Kaufmann, San Francisco (2005)
5. Baader, F., Küsters, R., Molitor, R.: Computing least common subsumers in description logics with existential restrictions. In: Proc. IJCAI 1999. Morgan Kaufmann, San Francisco (2003)
6. Brandt, S.: Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In: Proc. ECAI 2004. IOS Press, Amsterdam (2004)
7. Brandt, S.: Standard and Non-standard reasoning in Description Logics. Ph.D. dissertation, Institute for Theoretical Computer Science, TU Dresden, Germany (2006)
8. Brandt, S., Model, J.: Subsumption in $\mathcal{EL}$ w.r.t. hybrid TBoxes. In: Furbach, U. (ed.) KI 2005. LNCS (LNAI), vol. 3698. Springer, Heidelberg (2005)
9. Hofmann, M.: Proof-theoretic approach to description-logic. In: Proc. LICS 2005. IEEE Press, Los Alamitos (2005)
10. Model, J.: Subsumtion in $\mathcal{EL}$ bezüglich hybrider TBoxen. Diploma thesis, Institute for Theoretical Computer Science, TU Dresden, Germany (2005)
11. Novakovic, N.: Proof-theoretic approach to subsumption and least common subsumer in $\mathcal{EL}$ w.r.t. hybrid TBoxes. Master thesis, Institute for Theoretical Computer Science, TU Dresden, Germany (2007)
12. Rector, A., Horrocks, I.: Experience building a large, re-usable medical ontology using a description logic with transitivity and concept inclusions. In: Proc. AAAI 1997. AAAI Press, Menlo Park (1997)
13. Baader, F.: Computing the least common subsumer in the description logic $\mathcal{EL}$ w.r.t. terminological cycles with descriptive semantics. In: Ganter, B., de Moor, A., Lex, W. (eds.) ICCS 2003. LNCS, vol. 2746, pp. 117–130. Springer, Heidelberg (2003)
14. Spackman, K.: Managing clinical terminology hierarchies using algorithmic calculation: Experience with SNOMED-RT. Journal of the American Medical Informatics Association (2000); Fall Symposium Special Issue
15. Baader, F., Novakovic, N., Suntisrivaraporn, B.: A Proof-Theoretic Subsumption Reasoner for Hybrid $\mathcal{EL}$-TBoxes. In: Proceedings of the 2008 International Workshop on Description Logics (DL 2008). CEUR-WS (2008)