Technical University Dresden
Department of Computer Science

Master's Thesis on

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

by

Novak Novaković

born on 2nd October 1981 in Belgrade

Supervising Professor: Prof. Dr. Franz Baader

Dresden, September 2007.

Technische Universität Dresden

Author	:	**Novak Novaković**
Matrikel-Nr.	:	**3177963**
Title	:	**Proof-theoretic Approach to Deciding Subsumption and Computing Least Common Subsumer in $\mathcal{EL}$ w.r.t. Hybrid TBoxes**
Degree	:	**Master of Science**
Date of submission	:	

Declaration

Hereby I certify that the thesis has been written by me. Any help that I have received in my research work has been acknowledged. Additionally, I certify that I have not used any auxiliary sources and literature except those cited in the thesis.

———————————————

Novak Novaković

Abstract

Description Logics (DLs) are logic-based formalisms for representing knowledge that provide a trade-off in expressivity and reasoning within a formalism. On one side of the spectrum, there is a light-weight description logic $\mathcal{EL}$ with top symbol, conjunction and existential restriction, whose expressivity, though very limited, is sufficient to describe some of the widely used ontologies of today. It has been shown that the subsumption problem in $\mathcal{EL}$ has polynomial complexity for terminologies with cyclic definitions w.r.t. greatest fixpoint, least fixpoint, and descriptive semantics, and for the general terminologies interpreted by descriptive semantics. Recently, the same was shown in these cases by devising sound and complete proof systems for these semantics.

Hybrid $\mathcal{EL}$ TBoxes have been proposed in order to combine the generality of cyclic definitions and general concept inclusions with certain desirable non-standard inference services. These terminologies combine two of the semantics, descriptive and greatest fixpoint semantics. A polynomial subsumption algorithm for theories with hybrid $\mathcal{EL}$ TBoxes and corresponding semantics was recently proposed.

Motivated by previous results, this thesis looks at the problem of subsumption w.r.t. hybrid $\mathcal{EL}$ TBoxes from a proof-theoretic point of view. A rule system is devised, and we show that polynomial proof search in the calculus yields a polynomial decision procedure for the subsumption problem in $\mathcal{EL}$ w.r.t. hybrid TBoxes.

In addition, we consider the problem of existence of least common subsumers of two $\mathcal{EL}$ concepts w.r.t. hybrid TBoxes. We give a positive answer to the existence problem by providing an algorithm for computing such concepts and showing its correctness in a proof-theoretic manner.

Acknowledgements

First of all, I am deeply grateful to Prof. Franz Baader for introducing me to scientific research in the field of description logics. It was great honor, a source of motivation and a pleasure to be supervised by him. Without his great support and infinite patience this thesis would not exist.

Also, I am grateful to Boontawee (Meng) Suntisrivaraporn for his suggestions and great support in implementing some of the algorithms presented in this thesis.

In addition, I want to thank Prof. Alexander Leitsch, Prof. Christian Fermüller, Prof. Mathias Baaz and Prof. Thomas Eiter for their inspired lectures in logic given last year and for their contagious enthusiasm. Many thanks go to Prof. Alexander Leitsch and Prof. Steffen Hölldobler, for their precious advice, and their constant support during my studying in the Computational Logic Programme.

I should also express my thankfulness to European Commission for their financial support through the Erasmus Mundus Programme. Their grant enabled me to study in the International Masters Programme in Computational Logic.

Special thanks go to my friends from Vienna and Dresden that made my life and studying logic in these two cities pure pleasure.

Most importantly, I am grateful to my parents and my brother for their endless love, support and joy.

Chapter 1

Introduction

Knowledge Representation (KR) is a major field of research within Artificial Intelligence. It investigates the approaches to store knowledge about a given domain of discourse in an explicit form in a knowledge base, and to automatically infer implicit consequences of the information stored. A KR system therefore consists of a knowledge base equipped with an algorithm for inferring implicit knowledge in domain-independent way.

One of the most prominent examples of such KR systems are Description Logics (DLs). Description logics ([11]) are a family of logic-based knowledge representation formalisms designed to represent and reason about conceptual knowledge in a structured and semantically well-understood way. They are a family of formal languages that provide specialized inference mechanisms to account for knowledge classification, while preserving decidability and allowing for the design of efficient reasoning algorithms. Knowledge is represented in terms of concepts (i.e., classes of objects) and roles (i.e., binary relationships between classes).

Description Logics originate form early KR systems, such as frame-based systems and semantic networks ([14]). These early formalisms were not equipped with a formal, logic-based semantics, and the development of a formal semantics led to the notion of DLs. It was shown later that some DLs are in a close correspondence with representation formalisms used in databases and software engineering (see Chapter "Relationship with other formalisms" from [11]). Ever since, there is an increasing number of applications where DLs play important role, e.g., in the foundation of logical ontology languages which are used in several areas such as biomedical ontologies, and natural language processing.

As fragments of first-order logic, DLs sacrifice some of the expressivity of the entire first-order logic in favor of decidability of reasoning tasks. There is a clear tradeoff between the complexity of a reasoning task in a description logic and the expressivity of the logic. On one side of this spectrum there are very expressive description logics like the ones from the $\mathcal{SHOIQ}$ family where some of the reasoning tasks are $NExpTime$-hard, if not undecidable. In contrast, description logics from the $\mathcal{EL}$ family are light-weight logics with very restricted

expressivity. Despite its limited expressivity, some of important ontologies or its significant parts, such as GALEN ([24], [27]), SNOMED ([25], [29]) and Gene Ontology ([26], [1]) can be described in $\mathcal{EL}$ and its extensions. Moreover, much of the reasoning tasks turn out to be tractable for $\mathcal{EL}$, making this logic interesting for applications. These are the reasons for detailed investigation of this logic, and this work is another step in that direction.

1.1 Subsumption in $\mathcal{EL}$ - previous results

In [10], Baader has shown that the subsumption problem in $\mathcal{EL}$ has polynomial complexity for terminologies with cyclic definitions w.r.t. greatest fixpoint, least fixpoint, and descriptive semantics (as introduced by B. Nebel [22]). The main tool for the investigation of cyclic definitions in $\mathcal{EL}$ is a characterization of subsumption through the existence of a so-called simulation relation on the graph associated with an $\mathcal{EL}$ terminology, so called description graph. This relation on graphs can be computed in polynomial time [19]. For descriptive subsumption, a notion of synchronized simulation is introduced on the associated graphs, and subsumption w.r.t. descriptive semantics is associated with the existence of synchronized simulation between two graphs.

In [17] and [3], Brandt, Baader and Lutz showed that the subsumption problem in $\mathcal{EL}$ has polynomial complexity for terminologies with GCIs interpreted by descriptive semantics. This time, the polynomial decision procedure relays on so-called implication sets assigned to each concept. These sets are created in polynomial time by exhaustive application of fixed extension rules on the initial sets. A concept is subsumed by another one iff the later one belongs to the implication set of the former one.

Hofmann [20] considers the problem of subsumption in $\mathcal{EL}$ from a different perspective. He introduces a sound and complete calculus for providing proofs of subsumption relationships. Two calculi are devised, depending on wether a subsumption is decided modulo descriptive semantics with GCIs or the greatest fixpoint semantics w.r.t. cyclic definitions. A subsumption holds iff there is a proof in the corresponding calculus. The decision procedure for subsumption problem identifies all provable subsumptions in polynomial time. It is interesting that unlike the simulation approach, subsumption w.r.t. descriptive semantics seems to be easier to show, since it is similar to standard Genzen's sequent calculus with equality and the CUT rule is admissible. In the case of descriptive semantics with GCIs, soundness of the calculus is trivial and it can be proven using induction on derivation. Completeness of the calculus in this setting means that for every subsumption that follows from the TBox there is a proof for it. The main tool for showing completeness is 'universal interpretation' where a concept is interpreted by the set of provably subsumed concept descriptions. For the cyclic TBoxes interpreted by greatest fixpoint semantics, the notion of a proof of a subsumption is different. A subsumption is said to be proven if there exists certain infinite sequence of derivations rather than a single one.

As noted in [18] by Brandt and Model, there are two somewhat incompatible,

but desirable, features of DL systems: the support of general TBoxes containing general concept inclusion axioms, and non-standard inference services facilitating knowledge engineering tasks. It has been pointed out in [4] that some of the non-standard inference services, namely least-common subsumer and most specific concept, facilitate 'bottom-up' construction of DL terminologies. Here 'bottom-up' building of DL terminologies means beginning of building a terminology by selecting a set of example instances and using them to construct a new concept description intended to represent them. Having general concept inclusions in a TBox requires the TBox to be interpreted with descriptive semantics. In the same time it has been proved in [9] that the least-common subsumer and most specific concept need not exist when a TBox is interpreted w.r.t. descriptive semantics. That was the main motivation for Brandt and Model to propose *hybrid TBoxes* for $\mathcal{EL}$.

A hybrid $\mathcal{EL}$ TBox is a pair $(\mathcal{F}, \mathcal{T})$ consisting of a general TBox $\mathcal{F}$ and a (possibly) cyclic TBox $\mathcal{T}$ defined over the same set of atomic concepts and roles. $\mathcal{F}$ serves as a foundation of $\mathcal{T}$ in the sense that the GCI's of $\mathcal{F}$ define relationships between concepts used as primitive concept names in the definitions of $\mathcal{T}$, i.e. $\mathcal{F}$ 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 GCI's in $\mathcal{F}$. Hence, the foundation is interpreted by descriptive semantics, while the terminology is interpreted by greatest fixpoint semantics.

Brandt and Model also provide a decision procedure for the subsumption problem w.r.t. the semantics defined for hybrid $\mathcal{EL}$ TBoxes. They introduce the notion of $\mathcal{F}$-completion that denotes a polynomial reduction of an instance of the subsumption problem w.r.t. a hybrid $\mathcal{EL}$ TBox $(\mathcal{F}, \mathcal{T})$ to an instance of the subsumption problem w.r.t. the changed theory $f(\mathcal{T})$ interpreted by greatest fixpoint semantics. The reduction involves interpreting the hybrid TBox by descriptive semantics, i.e., treating $\mathcal{F} \cup \mathcal{T}$ as a set of GCIs interpreted by descriptive semantics. Then, all descriptive implications in $\mathcal{T}$ directly involving names from $\mathcal{F}$ are added to the definitions in $\mathcal{T}$ in order to obtain $f(\mathcal{T})$.

As mentioned before, the least common subsumer of two defined concepts is a non-standard inference service that facilitates 'bottom-up' building of DL terminologies. For two given defined concepts from an $\mathcal{EL}$ TBox, their least common subsumer is, as suggested by its name, a concept that subsumes the both defined concepts, and has the property of minimality, i.e., every other concept that subsumes the both of the defined concepts also subsumes their least common subsumer. The least common subsumer is defined in what is called a conservative extension of the TBox, and it enjoys the property of minimality in other conservative extensions of the original one. This obviously extends the syntax of the original TBox and makes the search for the least common subsumer more difficult. In fact, as mentioned before, it has been proven in [9] that least common subsumer need not exist if we interpret $\mathcal{EL}$ TBoxes with descriptive semantics. An algorithm for computing least common subsumer of two defined concepts from a $\mathcal{EL}$ TBox with cyclic definitions interpreted with greatest fixpoint semantics is shown by Baader in [8]. A correspondence between

normalized TBoxes with cyclic definitions and the assigned description graphs was exploited to obtain explicit definitions of least common subsumers from the so called product description graph. Starting from the TBox, the corresponding description graph $\mathcal{G}$ is constructed, and so is the product graph $\mathcal{G} \times \mathcal{G}$. Now, the conservative extension is obtained from the inverse mapping to the one that maps TBoxes into description graphs.

Following the idea of reducing subsumption problem w.r.t. hybrid $\mathcal{EL}$ TBoxes, to the one for $\mathcal{F}$-completions interpreted by greatest fixpoint semantics, Brandt shows in [15] that least common subsumers of two defined concepts from a hybrid TBox always exist and can be computed from the product description graph of $\mathcal{F}$-completion of the original TBox. Again, two different kinds of reasoning, the one for the descriptive semantics and the one for the greatest fixpoint semantics are serialized and the former always precedes the later in order to perform reasoning in the hybrid case.

1.2 Objectives and structure of the thesis

This paper presents an alternative polynomial decision procedure for the subsumption problem in hybrid $\mathcal{EL}$ TBoxes that is motivated by the proof-theoretic approach from [20]. This time, rather than reducing the reasoning in a hybrid TBox to the reasoning with greatest fixpoint semantics, we try to combine the two existing reasoning techniques into a single one. Again, a rule system is devised; a notion of the proof is introduced; and soundness and completeness are proven for the obtained calculus, i.e. we show that every subsumption that follows form the TBox can be derived in the devised calculus, and vice versa, every derivable subsumption holds w.r.t. the TBox. Since both $\mathcal{EL}$ TBoxes consisting of GCIs interpreted by descriptive semantics, and $\mathcal{EL}$ TBoxes with cyclic definitions interpreted by greatest fixpoint semantics, are special cases of hybrid $\mathcal{EL}$ TBoxes with the corresponding semantics, one can ask if the resulting rule system is a fusion of the two systems, the one for GCIs and the one for cyclic terminologies. This question turns out to be nontrivial, and we give a positive answer to it in this thesis.

Different rule systems yield different characteristics of the calculus, and thus different proofs of soundness and completeness are required.

In addition, we present a polynomial decision procedure for the subsumption problem w.r.t. hybrid TBoxes based on such a calculus.

The second main task solved in this thesis is to develop an algorithm for computing the least common subsumer of two defined concepts from a hybrid $\mathcal{EL}$ TBox. To that purpose, we employ the the developed proof-theoretic technique.

The rest of this thesis is organized as follows:

In Chapter 2 we introduce relevant definitions of the description logic $\mathcal{EL}$, its syntax and two types of semantics, descriptive and greatest fixpoint. The

notion of a hybrid TBox is introduced as well as the corresponding semantics. We also briefly introduce to inference problems w.r.t. hybrid TBoxes.

In Chapter 3 we devise a rule system for deciding subsumption w.r.t hybrid $\mathcal{EL}$ TBoxes. The notion of a proof is introduced, some of the properties of the system are analyzed, and subsequently, soundness and completeness are shown. In addition, a polynomial time algorithm for deciding subsumption based on the rule system is given.

Chapter 4 is dedicated to computation of the least common subsumer of two given concepts from a hybrid $\mathcal{EL}$ TBox. A construction algorithm is given and its correctness proven.

In Chapter 5 we derive conclusions, take another look at the existing results, and comment on the practical performance of the implemented system for deciding subsumption w.r.t. hybrid $\mathcal{EL}$ based on our decision procedure. We also consider other related topics and discuss possible future work.

Chapter 2

Preliminaries

In this chapter, we introduce notions and give formal definitions relevant for further discussion. Among other things, we present the description logic $\mathcal{EL}$, its syntax and semantics. Two types of semantics are discussed, depending on DL knowledge bases used, and hybrid TBoxes introduced. Two definitions of semantics for hybrid TBoxes are introduced and its equivalence proven. Finally, we introduce inference problem of deciding subsumption w.r.t. semantics for hybrid TBoxes.

2.1   Description logic $\mathcal{EL}$

Description logics provide a logical basis for knowledge representation. As such, they use formal syntax to denote objects of interest (or *individuals*), classes of objects (or *concepts*), and relationships between objects (or *roles*). As a particular description logic, $\mathcal{EL}$ is characterized by the set of *constructors* with which complex *concept descriptions* can be built from *atomic concept* names and *role names*.

N_{prim}, N_{role} and N_{def} will denote sets of *primitive concept names*, *role names*, and *defined concept names*, respectively. We denote elements of these sets by $P, Q, R, ...$; $r, s, t, ...$; and $X, Y, Z, ...$, respectively. Sometimes we may use indexes.

Complex concept descriptions in $\mathcal{EL}$ are formed using the constructors top-concept: ($\top$), conjunction ($C \sqcap D$) and existential restriction ($\exists r.C$). Formally,

Definition 2.1.1. *($\mathcal{EL}$ syntax)* Let N_{prim}, N_{def} and N_{role} be disjoint sets of primitive concept names, defined concept names and role names. The set of $\mathcal{EL}$-concept descriptions (or $\mathcal{EL}$-concepts) is the smallest set that is inductively defined as follows:

- each $P \in N_{prim}$ is an (primitive, atomic) $\mathcal{EL}$-concept;

- each $X \in N_{def}$ is an (defined, atomic) $\mathcal{EL}$-concept;

- if ϕ, ψ are $\mathcal{EL}$-concepts and $r \in N_{role}$ is a role name, then the top-concept $\top$, the conjunction $\phi \sqcap \psi$ and the existential restriction $\exists r.\phi$ are also $\mathcal{EL}$-concepts.

Greek letters $\alpha, \beta, \phi, \psi, \theta, ...$ will be used to denote arbitrary concept descriptions.

Definition 2.1.2. *($\mathcal{EL}$ semantics)* The semantics of $\mathcal{EL}$ concept descriptions is defined in terms of *interpretation* $\mathcal{I} = (\mathcal{D}_\mathcal{I}, \mathcal{I}(\cdot))$. The domain $\mathcal{D}_\mathcal{I}$ of $\mathcal{I}$ is a non-empty set of individuals, and the interpretation function $\mathcal{I}(\cdot)$ maps each concept name $A \in N_{prim} \cup N_{def}$ to a subset of $\mathcal{D}_\mathcal{I}$ and each role $r \in N_{role}$ to a binary relation $\mathcal{I}(r)$ on $\mathcal{D}_\mathcal{I}$. The extension of $\mathcal{I}(\cdot)$ to arbitrary concept descriptions is given as follows:

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

A concept C is *satisfiable* iff there is an interpretation $\mathcal{I}$ such that $\mathcal{I}(C) \neq \emptyset$. In this case, we say that $\mathcal{I}$ is a model of C. A concept C is *subsumed* by a concept D (written $C \sqsubseteq D$) iff for every interpretation $\mathcal{I}$, $\mathcal{I}(C) \subseteq \mathcal{I}(D)$. Two concepts C, D are equivalent (written $C \equiv D$) iff $C \sqsubseteq D$ and $D \sqsubseteq C$.

Sometimes, instead of writing $\mathcal{I}(\phi)$, we will write $\phi^\mathcal{I}$ for readability reasons.

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

Table 2.1: Summary - syntax and semantics of $\mathcal{EL}$

As an knowledge representation formalism, DLs facilitate intensional representation of the domain of interest through usage of certain hierarchical structures called terminologies.

Definition 2.1.3. *(TBox)* A *terminological axiom* is a concept definition or a general concept inclusion (GCI). A *concept definition* is an expression of the form $X \equiv \phi_X$ where X is a concept name from the set N_{def} and ϕ_X is a concept description. A general concept inclusion is an expression of the form $\phi \sqsubseteq \psi$, where ϕ, ψ are concept description. A TBox is a finite set of concept definitions, while general TBox is a finite set of GCIs.

Let $\mathcal{T}$ be a TBox that contains only concept definitions. Concept names occurring on the left-hand side of a definition are defined concepts. All other concept names occurring in the TBox are primitive concepts.

We say that $\mathcal{T}$ contains multiple definitions iff there are two distinct concepts ϕ_1 and ϕ_2 such that both $X \equiv \phi_1$ and $X \equiv \phi_2$ belong to $\mathcal{T}$. We also say that $\mathcal{T}$ contains a terminological cycle iff there is a subset $\{X_1 \equiv \phi_{X_1}, ..., X_n \equiv \phi_{X_1}\} \subseteq \mathcal{T}$ such that

- X_{i+1} appears in ϕ_{X_i}, for $1 \leq i \leq n$, and

- X_1 appears in ϕ_{X_n}.

$\mathcal{T}$ is called an acyclic TBox iff it contains no multiple definition and no terminological cycle.

We say that an interpretation $\mathcal{I}$ is a *(descriptive) model* of the TBox $\mathcal{T}$ iff it satisfies all its terminological axioms, i.e., $\mathcal{I}(X) = \mathcal{I}(\phi_X)$ for every $X = \phi_X$ in $\mathcal{T}$, and $\mathcal{I}(\phi) \subseteq \mathcal{I}(\psi)$ for every $\phi \sqsubseteq \psi$ in $\mathcal{T}$.

A concept ϕ is satisfiable w.r.t. a TBox $\mathcal{T}$ iff there is a model $\mathcal{I}$ of $\mathcal{T}$ such that $\mathcal{I}(C) \neq \emptyset$. A concept ϕ is subsumed by a concept ψ w.r.t. a TBox $\mathcal{T}$, denoted by $\mathcal{T} \models \phi \sqsubseteq \psi$ iff $\mathcal{I}(\phi) \subseteq \mathcal{I}(\psi)$ for every model $\mathcal{I}$ of $\mathcal{T}$. Two concepts ϕ, ψ are equivalent w.r.t. a TBox $\mathcal{T}$, denoted by $\mathcal{T} \models \phi \equiv \psi$ iff $\mathcal{I}(\phi) = \mathcal{I}(\psi)$ for every model $\mathcal{I}$ of $\mathcal{T}$.

The definition of general TBoxes given above is to general for our purposes. In the rest of this thesis we will restrict our attention to the TBoxes without multiple definitions. However, we will still allow for terminological cycles, i.e. we will assume that our TBoxes may contain cyclic definitions.

The semantics given above has been introduced as *descriptive semantics* by Nebel, [22]. In that context, we speak also of *descriptive subsumption* and *descriptive equivalence*.

Definition 2.1.4. *(ABox)* Let N_I be a set of individual names. An assertion axiom is an expression of the form $C(x)$ (called concept assertion) or $r(x, y)$ (called role assertion), where $x, y \in N_I$ are individual names, C a concept and r a role. A concept assertion is called simple if it is of the form $A(x)$ where A is a concept name. A role assertion is called simple whenever it is of the form $r(x, y)$ with r a role name, i.e., not a complex role expressions. An ABox $\mathcal{A}$ is a finite set of assertion axioms. Here, $\mathcal{A}$ is called simple whenever all of its assertions are simple. For the semantics, we require every interpretation additionally, to map each individual name $x \in N_I$ to an element $\mathcal{I}(x) \in \mathcal{D}_\mathcal{I}$. An interpretation $\mathcal{I}$ satisfies a concept assertion $C(x)$ iff $\mathcal{I}(x) \in \mathcal{I}(C)$ and it satisfies a role assertion $r(x, y)$ iff $(\mathcal{I}(x), \mathcal{I}(y)) \in \mathcal{I}(r)$. It satisfies an ABox $\mathcal{A}$ iff it satisfies every assertion axiom in $\mathcal{A}$. If such an interpretation $\mathcal{I}$ exists, then we say that $\mathcal{A}$ is satisfiable and we say that $\mathcal{I}$ is model of $\mathcal{A}$.

Definition 2.1.5. *(Knowledge base)* A knowledge base (KB) $\mathcal{K} = (\mathcal{T}, \mathcal{A})$ consists of a TBox $\mathcal{T}$ and an ABox $\mathcal{A}$. An interpretation $\mathcal{I}$ satisfies $\mathcal{K}$ iff $\mathcal{I}$ is a model of both $\mathcal{T}$ and $\mathcal{A}$. In this case, $\mathcal{K}$ is satisfiable and we say that $\mathcal{I}$ is a

model of $\mathcal{K}$. A concept ϕ is satisfiable w.r.t. a knowledge base $\mathcal{K}$ iff there is a model $\mathcal{I}$ of $\mathcal{K}$ such that $\mathcal{I}(\phi) \neq \emptyset$. A concept ϕ is subsumed by a concept ψ w.r.t. a knowledge base $\mathcal{K}$ iff $\mathcal{I}(C) \subseteq \mathcal{I}(D)$ for every model $\mathcal{I}$ of $\mathcal{K}$. Two concepts ϕ, ψ are equivalent w.r.t. a knowledge base $\mathcal{K}$ iff $\mathcal{I}(C) = \mathcal{I}(D)$ for every model $\mathcal{I}$ of $\mathcal{K}$.

Definition of ABoxes given above is introduced for the reasons of completeness of discussion on description logic $\mathcal{EL}$. In the rest of this thesis we will restrict our attention to reasoning problems w.r.t. TBoxes alone, i.e. we will always assume that our knowledge base will have empty ABox.

Before proceeding with greatest fixpoint semantics and introducing hybrid TBoxes, we define notion of subconcept of a concept description. This notion will be frequently used in the later chapters. This is not surprising, since the key to most of the proof-theoretic analysis is syntactic structure of the formulae of discourse.

Definition 2.1.6. Let ϕ be a concept description. Then the set of all *subconcepts* of concept description ϕ is the least set $SC(\phi)$such that

1. $\phi \in SC(\phi)$

2. if ϕ is of the form $\phi_1 \sqcap \phi_2$, then $SC(\phi) \supseteq SC(\phi_1) \cup SC(\phi_2)$

3. if ϕ is of the form $\exists r.\phi_1$, then $SC(\phi_1) \subseteq SC(\phi)$.

In addition, we say that ψ is a subconcept that occurs in a TBox $\mathcal{T}$ if $\mathcal{T}$ contains a definition $X \equiv \phi_X$ and ψ is a subconcept of X or ϕ_X, or if $\mathcal{T}$ contains a GCI $\theta \sqsubseteq \rho$ and ψ is a subconcept of θ or ρ.

2.2 Greatest fixpoint semantics. Hybrid TBoxes.

In previous section we introduced descriptive semantics for $\mathcal{EL}$ TBoxes. Defining greatest fixpoint (gfp) semantics requires certain preparation. To that purpose, we recall some of the definitions form [9] and [18].

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

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

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

$$\mathcal{I}_1 \preceq_{\mathcal{J}} \mathcal{I}_2 \text{ iff } \mathcal{I}_1(X) \subseteq \mathcal{I}_2(X) \text{ for all } X \in N_{def}.$$

Ordered pair $(Int(\mathcal{J}), \preceq_{\mathcal{J}})$ is a complete lattice on $Int(\mathcal{J})$. Indeed, it is easy to see that every subset of $Int(\mathcal{J})$ has a least upper bound (lub) and a greatest lower bound (glb). Thus, Tarski's fixpoint theorem [32] can be applied to this lattice and it will yield the claim that every monotonic function on $Int(\mathcal{J})$ has a fixpoint.

Definition 2.2.2. Let 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.

$$O_{\mathcal{T},\mathcal{J}} : Int(\mathcal{J}) \to Int(\mathcal{J}); \quad \mathcal{I}_1 \mapsto \mathcal{I}_2 \text{ iff } \mathcal{I}_2(X) = \mathcal{I}_1(\phi_X)$$

for all $X \equiv \phi_X \in \mathcal{T}$.

It was shown in [9] that $O_{\mathcal{T},\mathcal{J}}$ is in fact monotonous and can be used as a fixpoint operator on $Int(\mathcal{J})$. As a result, we obtain the following proposition.

Proposition 2.2.1. *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 $\mathcal{T}$.*

Previous proposition allows ua to define notion of fixpoint models for $\mathcal{EL}$-TBoxes as follows.

Definition 2.2.3. *(Greatest fixpoint semantics)* 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 $\mathcal{I} \in Int(\mathcal{J})$ is the greatest fixpoint of $O_{\mathcal{T},\mathcal{J}}$. Greatest fixpoint semantics considers only gfp-models as admissible models.

As a complete lattice, $(Int(\mathcal{J}), \preceq_{\mathcal{J}})$ uniquely determines the gfp-model for a given TBox $\mathcal{T}$ and a primitive interpretation $\mathcal{J}$. Thus we refer to the gfp-model $\mathrm{gfp}(\mathcal{T}, \mathcal{J})$ for any given $\mathcal{T}$ and $\mathcal{J}$.

Computing gfp-model $gfp(\mathcal{T}, \mathcal{J})$ requires defining iteration of $O_{\mathcal{T},\mathcal{J}}$ over ordinals.

Definition 2.2.4. 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 $\mathcal{I}_{\mathcal{J}}^{top} \in Int(\mathcal{J})$ by setting $\mathcal{I}_{\mathcal{J}}^{top}(X) := \mathcal{D}_{\mathcal{J}}$ for every $X \in N_{def}$. For every ordinal α, define

- $\mathcal{I}_{(\mathcal{T},\mathcal{J})}^{\downarrow\alpha} := \mathcal{I}_{\mathcal{J}}^{top}$ if $\alpha = 0$;

- $\mathcal{I}_{(\mathcal{T},\mathcal{J})}^{\downarrow\alpha+1} := O_{\mathcal{T},\mathcal{J}}(\mathcal{I}^{\downarrow\alpha})_{(\mathcal{T},\mathcal{J})}$;

- $\mathcal{I}_{(\mathcal{T},\mathcal{J})}^{\downarrow\alpha} := glb(\{\mathcal{I}_{(\mathcal{T},\mathcal{J})}^{\downarrow\beta} \mid \beta < \alpha\})$ if α is a limit ordinal; here, $\}\updownarrow\lfloor$ stands for greatest lower bound.

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

14

Corollary 2.2.1. *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 $gfp(\mathcal{T}, \mathcal{J}) = I^{\downarrow\alpha}_{(\mathcal{T}, \mathcal{J})}$.*

If α is a limit ordinal then $\mathcal{I}^{\downarrow\alpha}_{(\mathcal{T}, \mathcal{J})}$ equals $\bigcap_{\beta < \alpha} \mathcal{I}^{\downarrow\beta}_{(\mathcal{T}, \mathcal{J})}$. Now we are ready to introduce gfp-subsumption.

Definition 2.2.5. Let $\mathcal{T}$ be an $\mathcal{EL}$-TBox and let $\phi, \psi \in N^{\mathcal{T}}_{def}$. Then, ϕ is subsumed by ψ w.r.t. gfp-semantics ($\phi \sqsubseteq_{gfp,} \psi$) iff $\mathcal{I}(\phi) \subseteq \mathcal{I}(\psi)$ holds for all gfp-models $\mathcal{I}$ of $\mathcal{T}$.

An alternative definition of greatest fixpoint semantics is given in [20]. The alternative version provides slightly more compact way to check for gfp models of an $\mathcal{EL}$ TBox.

Definition 2.2.6. (Greatest fixpoint semantics) Consider a TBox with, possibly circular, definitions only of the form $X \equiv \phi_X$. A model of such a TBox under the greatest fixpoint semantics is its descriptive model $\mathcal{I}$ with a domain $\mathcal{D}_{\mathcal{I}}$, which has the further property that whenever $\mathcal{J}$ is a function mapping concept descriptions over N_{prim}, N_{def} and N_{role} to subsets of $\mathcal{D}_{\mathcal{I}}$ in such way that

$$
\begin{array}{rcll}
\mathcal{J}(\top) & = & \mathcal{I}(\top) & (1) \\
\mathcal{J}(P) & = & \mathcal{I}(P) & (2) \\
\mathcal{J}(X) & \subseteq & \mathcal{J}(\phi_X) & (3) \\
\mathcal{J}(\phi \sqcap \psi) & = & \mathcal{J}(\phi) \cap \mathcal{J}(\psi) & (4) \\
\mathcal{J}(\exists r.\phi) & = & \{x \mid \exists y \in \mathcal{J}(\phi).\mathcal{I}(r)(x,y)\} & (5)
\end{array}
$$

then $\mathcal{J}(\phi) \subseteq \mathcal{I}(\phi)$ for all concepts ϕ.

What should be mentioned here is that the interpretation $\mathcal{I}$ form the definition above is unique for a fixed domain and fixed interpretation of primitive concepts. Even more, $\mathcal{I}$ can be computed from its restriction to primitive concepts and role names, as the union of all functions $\mathcal{J}$ satisfying (1) - (5). We exploit these conclusion to show equivalence of the two definitions of gfp semantics.

Proposition 2.2.2. *Definitions Definition 2.2.3 and Definition 2.2.6 are equivalent.*

Proof. We show equivalence of the two definitions by showing that they describe the same models once we fix the domain and the primitive interpretation. First we fix a domain $\mathcal{D}_{\mathcal{I}}$ and a primitive interpretation $\mathcal{J}$. Then we consider a gfp model $\hat{\mathcal{I}}$ with a domain $\mathcal{D}_{\mathcal{I}}$ of a given TBox $\mathcal{T}$ form the Definition 2.2.3, such that $\hat{\mathcal{I}} \in Int(\mathcal{J})$. We proceed by showing that there is unique interpretation $\mathcal{I}$ from Definition 2.2.6 with a domain $\mathcal{D}_{\mathcal{I}}$ such that $\mathcal{I} \in Int(\mathcal{J})$. Finally we prove $\hat{\mathcal{I}} = \mathcal{I}$.

Let $\mathcal{I} \in Int(\mathcal{J})$ be, as announced, the interpretation from Definition 2.2.6, let $\mathbb{S}$ denote the set of all functions from the same definition that satisfy (1)-(5), and let
$$\mathcal{L}(\phi) = \bigcup_{\mathcal{K} \in \mathbb{S}} \mathcal{K}(\phi).$$
By induction on structure of ϕ, we show that $\mathcal{L}(\phi) = \mathcal{I}(\phi)$

- $\phi = \top$: then $\mathcal{K}(\top) = \mathcal{L}(\top) = \mathcal{I}(\top) = \mathcal{D}_\mathcal{I}$.
- $\phi = P \in N_{prim}$: then $\mathcal{K}(P) = \mathcal{J}(P) = \mathcal{I}(P)$ for all $\mathcal{K} \in \mathbb{S}$, thus $\mathcal{L}(P) = \bigcup_{\mathcal{K} \in \mathbb{S}} \mathcal{K}(P) = \mathcal{I}(P)$.
- $\phi = X \in N_{def}$: then $\mathcal{K}(X) \subseteq \mathcal{I}(X)$ for all $\mathcal{K} \in \mathbb{S}$, thus $\mathcal{L}(X) = \bigcup_{\mathcal{K} \in \mathbb{S}} \mathcal{K}(P) \subseteq \mathcal{I}(X)$. On the other hand, $\mathcal{I}$ obviously fulfills (1)-(4) form Definition 2.2.6, i.e. $\mathcal{I} \in \mathbb{S}$, thus $\mathcal{L}(X) \supseteq \mathcal{I}(X)$.
- $\phi = \phi_1 \sqcap \phi_2$: then $\mathcal{K}(\phi_1 \sqcap \phi_2) = \mathcal{K}(\phi_1) \sqcap \mathcal{K}(\phi_2)$, thus $\mathcal{L}(\phi_1 \sqcap \phi_2) = \bigcup_{\mathcal{K} \in \mathbb{S}} \mathcal{K}(\phi_1 \sqcap \phi_2) = \bigcup_{\mathcal{K} \in \mathbb{S}} \mathcal{K}(\phi_1) \sqcap \mathcal{K}(\phi_2) = \bigcup_{\mathcal{K} \in \mathbb{S}} \mathcal{K}(\phi_1) \sqcap \bigcup_{\mathcal{K} \in \mathbb{S}} \mathcal{K}(\phi_2) = \mathcal{L}(\phi_1) \sqcap \mathcal{L}(\phi_2)$. Induction hypothesis applied on both of the conjuncts yields $\mathcal{L}(\phi_1 \sqcap \phi_2) = \mathcal{I}(\phi_1) \sqcap \mathcal{I}(\phi_2) = \mathcal{L}(\phi_1 \sqcap \phi_2)$.
- $\phi = \exists r.\phi_1$: then $\mathcal{K}(\exists r.\phi_1) = \{x \mid \exists y \in \mathcal{K}(\phi_1).\mathcal{I}(r)(x,y)\}$, thus $\mathcal{L}(\exists r.\phi_1) = \bigcup_{\mathcal{K} \in \mathbb{S}} \{x \mid \exists y \in \mathcal{K}(\phi_1).\mathcal{I}(r)(x,y)\}$. The last is eqial to $\{x \mid \exists y \in \bigcup_{\mathcal{K} \in \mathbb{S}} \mathcal{K}(\phi_1).\mathcal{I}(r)(x,y)\} = \{x \mid \exists y \in \mathcal{L}(\phi_1).\mathcal{I}(r)(x,y)\}$. Induction hypothesis yields $\mathcal{L}(\exists r.\phi_1) = \{x \mid \exists y \in \mathcal{I}(\phi_1).\mathcal{I}(r)(x,y)\} = \mathcal{I}(\exists r.\phi_1)$.

As two interpretations, $\hat{\mathcal{I}}$ and $\mathcal{I}$ interpret $\top$ as entire domain, therefore $\hat{\mathcal{I}}$ fulfills condition (1) form Definition 2.2.6. From $\hat{\mathcal{I}} \in Int(\mathcal{J})$, we conclude $\hat{\mathcal{I}}(P) = \mathcal{J}(P) = \mathcal{I}(P)$, i.e. $\hat{\mathcal{I}}$ fulfills condition (2) form Definition 2.2.6. In a similar way, $\hat{\mathcal{I}}$ fulfills condition (3), while condition (4) follows form the fact that $\hat{\mathcal{I}}$ is an interpretation. Condition (5) is fulfilled by $\hat{\mathcal{I}}$, since $\hat{\mathcal{I}}(X) = \phi_X$ for every $X \in N_{def}$. Therefore, $\hat{\mathcal{I}}$ is in $\mathbb{S}$, and thus $\mathcal{L}(X) \supseteq \hat{\mathcal{I}}(X)$ for all $X \in N_{def}$. This implies $\mathcal{L} = \mathcal{I} \succeq_\mathcal{J} \hat{\mathcal{I}}$.

On the other hand, $\mathcal{I}$ is an interpretation for which it holds $O_{\mathcal{T},\mathcal{J}}(\mathcal{I}) = \mathcal{I}$, i.e. $\mathcal{I}$ is a fixpoint of the operator $O_{\mathcal{T},\mathcal{J}}$. Since $\mathcal{I} \succeq_\mathcal{J} \hat{\mathcal{I}}$, $\hat{\mathcal{I}}$ is the greatest fixpoint of operator $O_{\mathcal{T},\mathcal{J}}$, this is only possible if $\hat{\mathcal{I}}(X) = \mathcal{I}(X)$ for all $X \in N_{def}$. This proves the proposition, since $\hat{\mathcal{I}}$ and $\mathcal{I}$ already coincide on role names and primitive concepts. $\qquad\square$

Notice that Definition 2.2.6 requires that $\mathcal{J}$ maps all concept descriptions over N_{prim}, N_{def} and N_{role} to subsets of $\mathcal{D}_\mathcal{I}$. However, it suffice to restrict $\mathcal{J}$ to subconcepts occurring in the TBox, as suggested in the following proposition:

Proposition 2.2.3. *Consider a TBox $\mathcal{T}$ with, possibly circular, definitions only of the form $X \equiv \phi_X$ and its descriptive model $\mathcal{I}$ with a domain $\mathcal{D}_\mathcal{I}$. Then, $\mathcal{I}$ is a gfp model of the TBox iff whenever $\mathcal{J}$ is a function mapping subconcepts occurring in $\mathcal{T}$ to subsets of $\mathcal{D}_\mathcal{I}$, satisfying conditions (1)-(5) of Definition 2.2.6, then $\mathcal{J}(\phi) \subseteq \mathcal{I}(\phi)$ for all concepts ϕ.*

Proof. We begin the proof by imposing an assumption that w.l.o.g., all the elements of N_{prim}, N_{def} and N_{role} are occuring in $\mathcal{T}$.

If $\mathcal{I}$ is a gfp model of $\mathcal{T}$, then whenever $\mathcal{J}$ is a function mapping concept descriptions over N_{prim}, N_{def} and N_{role} to subsets of $\mathcal{D}_{\mathcal{I}}$, satisfying conditions (1)-(5) of Definition 2.2.6, then $\mathcal{J}(\phi) \subseteq \mathcal{I}(\phi)$ for all concepts ϕ. In particular, this will hold for all such functions $\mathcal{J}$ that map subconcepts occurring in $\mathcal{T}$ to subsets of $\mathcal{D}_{\mathcal{I}}$.

Suppose now that whenever $\mathcal{J}$ is a function mapping subconcepts occurring in $\mathcal{T}$ to subsets of $\mathcal{D}_{\mathcal{I}}$, satisfying conditions (1)-(5) of Definition 2.2.6, then $\mathcal{J}(\phi) \subseteq \mathcal{I}(\phi)$ for all concepts ϕ. Let $\mathcal{K}$ be an arbitrary function mapping concept descriptions over N_{prim}, N_{def} and N_{role} to subsets of $\mathcal{D}_{\mathcal{I}}$, satisfying conditions (1)-(5). We show by induction on ϕ that $\mathcal{K}(\phi) \subseteq \mathcal{I}(\phi)$ for all concepts ϕ.

- $\phi = \top$: by condition (1) of Definition 2.2.6, $\mathcal{K}(\top) = \mathcal{I}(\top) = \mathcal{D}_{\mathcal{I}}$;
- $\phi \in N_{prim}$: by condition (2) of Definition 2.2.6, $\mathcal{K}(\phi) = \mathcal{I}(\phi)$;
- $\phi \in N_{def}$: consider the function $\mathcal{J}$ obtained by restricting the $\mathcal{K}$ to the subconcepts occurring in $\mathcal{T}$. It is clear that $\mathcal{J}$ fulfills (1)-(5), and thus $\mathcal{K}(\phi) = \mathcal{J}(\phi) \subseteq \mathcal{I}(\phi)$;
- $\phi = \phi_1 \sqcap \phi_2$: by condition (4) $\mathcal{K}(\phi_1 \sqcap \phi_2) = \mathcal{K}(\phi_1) \cap \mathcal{K}(\phi_2)$, and induction hypothesis applied to $\mathcal{K}(\phi_1)$ and $\mathcal{K}(\phi_2)$ yields $\mathcal{K}(\phi_1 \sqcap \phi_2) = \mathcal{K}(\phi_1) \cap \mathcal{K}(\phi_2) \subseteq \mathcal{I}(\phi_1) \cap \mathcal{I}(\phi_2) = \mathcal{I}(\phi_1 \sqcap \phi_2)$;
- $\phi = \exists r.\phi_1$: by condition (5), $\mathcal{K}(\phi) = \{x \mid \exists y \in \mathcal{K}(\phi_1).\mathcal{I}(r)(x,y)\}$. Induction hypothesis applied to $\mathcal{K}(\phi_1)$ yields $\mathcal{K}(\phi) = \{x \mid \exists y \in \mathcal{K}(\phi_1).\mathcal{I}(r)(x,y)\} \subseteq \{x \mid \exists y \in \mathcal{I}(\phi_1).\mathcal{I}(r)(x,y)\} = \mathcal{I}(\phi)$.

$\square$

It can easily be seen that gfp semantics is coarser than descriptive semantics. In fact, 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. $\sqsubseteq_{gfp,\mathcal{T}}$. The converse, though, does not hold, e.g., if the TBox contains the definitions $X \equiv P \sqcap \exists r.X$ and $Y \equiv P \sqcap \exists r.Y$, then $X \sqsubseteq_{gfp,\mathcal{T}} Y$ holds, while $X \sqsubseteq_{\mathcal{T}} Y$ does not hold.

We introduce now the central notion of this thesis, hybrid $\mathcal{EL}$ TBoxes.

Definition 2.2.7. *(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 over N_{prim} and N_{role} (a finite set of GCI's over N_{prim} and N_{role}) and $\mathcal{T}$ is an $\mathcal{EL}$-TBox over N_{prim}, N_{role} and N_{def}.

Hence, $\mathcal{F}$ is a finite set of GCI's of the form $\phi_i \sqsubseteq \psi_i$, and $\mathcal{T}$ is a finite set of terminologies of the form $X_i \equiv \phi_{X_i}$. An example of a hybrid $\mathcal{EL}$ terminology on simplified medical knowledge base, taken form [18], is given in Table 2.2.

It defines concepts 'disease of the connective tissue', 'bacterial infection' and 'bacterial pericarditis' in a cyclic manner, while, 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.

$\mathcal{T}$:	ConnTissDisease	$\equiv$	Disease $\sqcap$ $\exists$acts_on.ConnTissue
	BactInfection	$\equiv$	Infection $\sqcap$ $\exists$causes.BactPericarditis
	BactPericarditis	$\equiv$	Inflammation $\sqcap$ $\exists$has_loc.Pericardium
			$\sqcap$ $\exists$caused_by.BactInfection
$\mathcal{F}$:	Disease $\sqcap$ $\exists$has_loc.ConnTissue	$\sqsubseteq$	$\exists$acts_on.ConnTissue
	Inflammation	$\sqsubseteq$	Disease
	Pericardium	$\sqsubseteq$	ConnTissue

Table 2.2: An example of a hybrid $\mathcal{EL}$ TBox

Hybrid $\mathcal{EL}$ TBoxes induce certain fixed semantics. This semantics is introduced in the following definition.

Definition 2.2.8. Let $(\mathcal{F}, \mathcal{T})$ be a hybrid $\mathcal{EL}$-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 $C^{\mathcal{J}} \subseteq D^{\mathcal{J}}$ 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 $\mathcal{T}$.

We conclude this chapter with definition of subsumption decision problem w.r.t. hybrid $\mathcal{EL}$ TBoxes.

Definition 2.2.9. 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 $(\mathcal{F}, \mathcal{T})$.

Solving this problem is the main problem of consideration of this thesis.

Chapter 3

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

This chapter is dedicated to construction of a decision procedure for subsumption problem in $\mathcal{EL}$ w.r.t. hybrid TBoxes. to that purpose, we devise a proof-system in Section 3.1 whose soundness and and completeness are shown in sections 3.2 and 3.3, respectively. Coincidence of subsumption relation and provability will lead to a tractable decision procedure in Section 3.4.

3.1 Calculus

Suppose that a hybrid $\mathcal{EL}$-TBox $(\mathcal{F}, \mathcal{T})$ is given, and one has to decide wether subsumption $A \sqsubseteq B$ follows from the given hybrid TBox. As mentioned before, a Gentzen style calculus will be introduced that will eventually lead to a decision procedure for subsumption problem w.r.t. hybrid $\mathcal{EL}$-TBoxes. To this purpose, we define relations $\sqsubseteq_n$ for every $n \geq 0$ on subconcepts occurring in $(\mathcal{F}, \mathcal{T})$. Therefore, consider the rule system HC (**H**ybrid $\mathcal{EL}$ Tbox **C**alculus), given in the Table 3.1.

Here, as mentioned, relation $\sqsubseteq_n$ is defined for $n \in \mathbb{N}$, and ϕ, ψ, ρ and θ will range over the subconcepts that occur in $(\mathcal{F}, \mathcal{T})$. We use notion of *sequent* to denote expressions of the form $\psi \sqsubseteq_n \phi$, where ϕ and ψ are some subconcepts occurring in $(\mathcal{F}, \mathcal{T})$, and we refer to them as the left and right-hand side of the sequent, respectively.

Derivations of sequents in HC are defined in a standard manner.

Definition 3.1.1. (Proof tree)

1. Every instance of the rules (Ax), (Top) and (Start) is a root of a one-element proof-tree.

2. Let
$$\frac{R}{S}$$

$$\phi \sqsubseteq_n \phi \qquad \text{(Ax)} \qquad\qquad \phi \sqsubseteq_n \top \qquad \text{(Top)}$$

$$\frac{\phi \sqsubseteq_n \rho}{\phi \sqcap \psi \sqsubseteq_n \rho} \qquad \text{(AndL1)} \qquad \frac{\psi \sqsubseteq_n \rho}{\phi \sqcap \psi \sqsubseteq_n \rho} \qquad \text{(AndL2)}$$

$$\frac{\phi \sqsubseteq_n \rho \quad \phi \sqsubseteq_n \theta}{\phi \sqsubseteq_n \rho \sqcap \theta} \qquad \text{(AndR)}$$

$$\frac{\phi \sqsubseteq_n \psi}{\exists r.\phi \sqsubseteq_n \exists r.\psi} \qquad\qquad \text{(Ex)}$$

$$\phi \sqsubseteq_0 \psi \qquad\qquad \text{(Start)}$$

$$\frac{\phi_X \sqsubseteq_n \psi}{X \sqsubseteq_n \psi} \qquad\qquad \text{(DefL)} \qquad \text{for } X \equiv \phi_X \in \mathcal{T}$$

$$\frac{\psi \sqsubseteq_n \phi_X}{\psi \sqsubseteq_{n+1} X} \qquad\qquad \text{(DefR)} \qquad \text{for } X \equiv \phi_X \in \mathcal{T}$$

$$\frac{\phi \sqsubseteq_n \alpha \quad \beta \sqsubseteq_n \psi}{\phi \sqsubseteq_n \psi} \qquad \text{(Concept)} \qquad \text{for } \alpha \sqsubseteq \beta \in \mathcal{F}$$

Table 3.1: Rule system HC

be an instance of some of the rules (AndL1), (AndL2), (Ex), (DefL) or (DefR). If there is a proof tree $\mathcal{P}$ with the root R, then

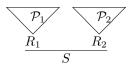$$\frac{\overline{\mathcal{P}}}{\dfrac{R}{S}}$$

is a proof tree with the root S.

3. Let

$$\frac{R_1 \quad R_2}{S}$$

be an instance of some of the rules (AndR) or (Concept). If there are two proof trees $\mathcal{P}_1$ and $\mathcal{P}_2$ with roots R_1 and R_2, respectively, then

$$\frac{\overline{\mathcal{P}_1} \qquad \overline{\mathcal{P}_2}}{\dfrac{R_1 \qquad R_2}{S}}$$

is a proof tree with the root S.

Definition 3.1.2. Sequent $\phi \sqsubseteq_n \psi$ is proven if there is a proof tree with the root $\phi \sqsubseteq_n \psi$. We say that ϕ and ψ are in relation $\sqsubseteq_n$, or we simply write $\phi \sqsubseteq_n \psi$.

Also, $\phi \sqsubseteq_\infty \psi$ will denote the claim that $\phi \sqsubseteq_n \psi$ can be derived for all $n \in \mathbb{N}$.

In Table 3.1 we recall the example of what could be a small real-life hybrid TBox. The derivation below demonstrates how BactPericarditis $\sqsubseteq_n$ ConnTissDisease can be derived for every n, therefore, how proof of BactPericarditis $\sqsubseteq_\infty$ ConnTissDisease can be obtained.

There are a few characteristics of the calculus HC that should be observed. First of all, one should notice that the given calculus considers only subconcepts occurring in the given hybrid $\mathcal{EL}$-TBox $(\mathcal{F}, \mathcal{T})$. It turns out that it is sufficient to consider only these concepts for the purposes of deciding subsumption. This fact will be shown in section on the completeness of the calculus, where we will refer to Proposition 2.2.3. Notice also that for the $\sqsubseteq_0$, every concept description is subsumed by every other concept description. As n increases, some of the concept descriptions stop being in $\sqsubseteq_n$ relation. One can also easily see that $\phi \sqsubseteq_n \psi$ implies $\phi \sqsubseteq_m \psi$ for all $m \leq n$ (proof goes by induction on derivation of $\phi \sqsubseteq_n \psi$), i.e. mapping $n \mapsto \sqsubseteq_n$ is monotone. Further more, $\sqsubseteq_n$ is defined on a finite set, hence there is a fixpoint relation $\sqsubseteq_{n_0}$, such that for all $n \in \mathbb{N}$ and all subconcepts ϕ and ψ holds: $n \geq n_0$ implies: ($\phi \sqsubseteq_n \psi$ iff $\phi \sqsubseteq_{n_0} \psi$). As a consequence, $\phi \sqsubseteq_\infty \psi$ is decidable, even more, it will be shown that this decision procedure is polynomial.

Before doing so, in the remaining part of this section, we analyze some of the properties of the proofs in HC.

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

$$
\cfrac{
a \quad \cfrac{
\cfrac{
\cfrac{
\cfrac{\cfrac{\text{Infl} \sqsubseteq_n \text{Infl} \; (Ax) \quad \text{D} \sqsubseteq_n \text{D} \; (Ax)}{\text{Infl} \sqsubseteq_n \text{D}} \; (Concept)}{\text{Infl} \sqcap \exists\text{hl.P} \sqsubseteq_n \text{D}} \; (AndL1) \quad b
}{\text{Infl} \sqcap \exists\text{hl.P} \sqsubseteq_n \text{D} \sqcap \exists\text{hl.CT}} \; (AndR) \quad \cfrac{\exists\text{ao.CT} \sqsubseteq_n \exists\text{ao.CT} \; (Ax)}{} \; (Concept)
}{\text{Infl} \sqcap \exists\text{hl.P} \sqsubseteq_n \exists\text{ao.CT}} \; (AndR)
}{\text{Infl} \sqcap \exists\text{hl.P} \sqsubseteq_n \text{D} \sqcap \exists\text{ao.CT}}
}{
\cfrac{
\cfrac{\text{Infl} \sqcap \exists\text{hl.P} \sqcap \exists\text{cb.BI} \sqsubseteq_n \text{D} \sqcap \exists\text{ao.CT}}{\text{BP} \sqsubseteq_n \text{D} \sqcap \exists\text{ao.CT}} \; (DefL)
}{\text{BP} \sqsubseteq_{n+1} \text{CTD}} \; (DefR)
} \; (AndL2)
$$

$$
a: \quad \cfrac{
\cfrac{\cfrac{\text{Infl} \sqsubseteq_n \text{Infl} \; (Ax) \quad \text{D} \sqsubseteq_n \text{D} \; (Ax)}{\text{Infl} \sqsubseteq_n \text{D}} \; (Concept)}{\text{Infl} \sqcap \exists\text{hl.P} \sqsubseteq_n \text{D}} \; (AndL1)
}{}
$$

$$
b: \quad \cfrac{
\cfrac{
\cfrac{\cfrac{\text{P} \sqsubseteq_n \text{P} \; (Ax) \quad \text{CT} \sqsubseteq_n \text{CT} \; (Ax)}{\text{P} \sqsubseteq_n \text{CT}} \; (Concept)}{\exists\text{hl.P} \sqsubseteq_n \exists\text{hl.CT}} \; (Ex)
}{\text{Infl} \sqcap \exists\text{hl.P} \sqsubseteq_n \exists\text{hl.CT}} \; (AndL2)
}{}
$$

Table 3.2: An example of a derivation in HC. Abbreviations used: CTD - ConnectionTissDisease, D - Disease, CT - ConnTissue, BI - BactInfection, BP - BactPericarditis, Infl - Inflammation, P - Pericardium, ao - acts_on, hl - has_location.

We begin by introducing an important relation $\Vdash$. This relation denotes derivations that only operate on the left-hand side of the relations $\phi \sqsubseteq \psi$. This rather technical relation will serve as an important tool in analyzing proofs in the defined calculus.

Definition 3.1.3. We write $\phi \Vdash \psi$ to denote that judgement $\phi \sqsubseteq_n \psi$ can be derived using rules (Ax), (AndL1), (AndL2), (DefL) and (Concept) for some n.

Notice that if $\phi \sqsubseteq_n \psi$ can be derived using rules (Ax), (AndL1), (AndL2), (DefL) and (Concept) for some n, then $\phi \sqsubseteq_n \psi$ for all $n \in \mathbb{N}$.
The following lemmas characterize, to a certain extent, properties of some of the proofs in the HC calculus.

Lemma 3.1.1. *If $\phi \Vdash \psi$ and $\psi \sqsubseteq_n \rho$ then $\phi \sqsubseteq_n \rho$.*

Proof. By induction on derivation of $\phi \Vdash \psi$.

- Base case: the last rule applied - (Ax) $\phi \Vdash \phi$
  here $\psi = \phi$, hence $\psi \sqsubseteq_n \rho$ implies $\phi \sqsubseteq_n \rho$
- Suppose the claim holds for all ϕ, ψ s.t. there is a derivation $\mathcal{P}$ of $\phi \Vdash \psi$ of the depth n or less.

  1. $\phi = \phi_1 \sqcap \phi_2$, and the last rule applied is (AndL1) (depth of $\mathcal{P}'$ is not greater that n):

  $$\frac{\overset{\displaystyle \mathcal{P}'}{\phi_1 \sqsubseteq_n \psi}}{\phi_1 \sqcap \phi_2 \sqsubseteq_n \psi} \ (AndL1)$$

  since $\psi \sqsubseteq_n \rho$, by the induction hypothesis (IH) $\phi_1 \sqsubseteq_n \rho$, hence $\phi \sqsubseteq_n \rho$ can be derived using (AndL1)

  2. $\phi = \phi_1 \sqcap \phi_2$, and the last rule applied is (AndL2) (depth of $\mathcal{P}'$ is not greater that n):

  $$\frac{\overset{\displaystyle \mathcal{P}'}{\phi_2 \sqsubseteq_n \psi}}{\phi_1 \sqcap \phi_2 \sqsubseteq_n \psi} \ (AndL2)$$

  since $\psi \sqsubseteq_n \rho$, by (IH) $\phi_2 \sqsubseteq_n \rho$, hence $\phi \sqsubseteq_n \rho$ can be derived using (AndL2)

  3. $\phi = X$ and the last rule applied is (DefL) (depth of $\mathcal{P}'$ is not greater that n):

  $$\frac{\overset{\displaystyle \mathcal{P}'}{X \sqsubseteq_n \psi}}{\phi_X \sqsubseteq_n \psi} \ (DefL)$$

  with $X \equiv \phi_X \in \mathcal{T}$. Since $\psi \sqsubseteq_n \rho$, by (IH) $\phi_X \sqsubseteq_n \rho$, hence $X \sqsubseteq_n \rho$ can be derived using (DefL)

4. the last rule applied is (Concept) (depths of $\mathcal{P}'$ and $\mathcal{P}''$ are not greater that n):

$$\frac{\phi \sqsubseteq_n \alpha \qquad \beta \sqsubseteq_n \psi}{\phi \sqsubseteq_n \psi} \ (Concept)$$

with $\alpha \sqsubseteq \beta \in \mathcal{F}$. Since $\psi \sqsubseteq_n \rho$, by (IH) $\beta \sqsubseteq_n \rho$, hence $\phi \sqsubseteq_n \rho$ can be derived using (Concept)

$\square$

Lemma 3.1.2. *Suppose that $n > 0$.*

1. $\theta \sqsubseteq_n \psi_1 \sqcap \psi_2$ iff $\theta \sqsubseteq_n \psi_1$ and $\theta \sqsubseteq_n \psi_2$,

2. $\theta \sqsubseteq_{n+1} X$ iff $\theta \sqsubseteq_n \phi_X$.

In addition, it will hold $\theta \sqsubseteq_\infty \psi_1 \sqcap \psi_2$ iff $\theta \sqsubseteq_\infty \psi_1$ and $\theta \sqsubseteq_\infty \psi_2$, and $\theta \sqsubseteq_\infty X$ iff $\theta \sqsubseteq_\infty \phi_X$.

Proof. The *if* directions are immediate using appropriate proof rules (AndR) and (DefR), respectively.

For the *only if* directions we proceed by induction on derivations. The idea is that the only rules applicable are among those rules that define $\Vdash$, where inductive argumentation comes through, or else the rule which decomposes the formula on the right hand side.

1. By induction on derivation of $\theta \sqsubseteq_n \psi_1 \sqcap \psi_2$.
   - Last rule used: (Ax) - induction base

     $$\psi_1 \sqcap \psi_2 \sqsubseteq_n \psi_1 \sqcap \psi_2 \text{ implies } \psi_1 \sqcap \psi_2 \sqsubseteq_n \psi_i, \text{ for } i = 1, 2$$

     usinf (Ax) and (AndLi)
   - Last rule used: (AndR)

     $$\frac{\theta \sqsubseteq_n \psi_1 \qquad \theta \sqsubseteq_n \psi_2}{\theta \sqsubseteq_n \psi_1 \sqcap \psi_2} \ (AndR)$$

     immediate
   - Last rule used: (AndLi) (for $i = 1, 2$)

     $$\frac{\theta_i \sqsubseteq_n \psi_1 \sqcap \psi_2}{\theta_1 \sqcap \theta_2 \sqsubseteq_n \psi_1 \sqcap \psi_2} \ (AndLi)$$

     (IH) yields $\theta_i \sqsubseteq_n \psi_1$ and $\theta_i \sqsubseteq_n \psi_2$, hence, using (AndLi) $\theta_1 \sqcap \theta_2 \sqsubseteq_n \psi_1$ and $\theta_1 \sqcap \theta_2 \sqsubseteq_n \psi_2$

24

- Last rule used: (DefL) ($\theta = X$)

$$\frac{\phi_X \sqsubseteq_n \psi_1 \sqcap \psi_2}{X \sqsubseteq_n \psi_1 \sqcap \psi_2} \ (DefL)$$

  (IH) yields $\phi_X \sqsubseteq_n \psi_1$ and $\phi_X \sqsubseteq_n \psi_2$, hence, using (DefL) $X \sqsubseteq_n \psi_1$ and $X \sqsubseteq_n \psi_2$

- Last rule used: (Concept)

$$\frac{\theta \sqsubseteq_n \alpha \quad \beta \sqsubseteq_n \psi_1 \sqcap \psi_2}{\theta \sqsubseteq_n \psi_1 \sqcap \psi_2} \ (Concept) \text{ with } \alpha \sqsubseteq \beta \in \mathcal{F}$$

  (IH) yields $\beta \sqsubseteq_n \psi_1$ and $\beta \sqsubseteq_n \psi_2$, hence, using (Concept) $\theta \sqsubseteq_n \psi_1$ and $\theta \sqsubseteq_n \psi_2$.

2. Analogous as for 1. :

- (Ax) $X \sqsubseteq_{n+1} X$

$$\frac{\phi_X \sqsubseteq_n \phi_X}{X \sqsubseteq_n \phi_X} \ (DefL)$$

- (DefR)

$$\frac{\theta \sqsubseteq_n \phi_X}{\theta \sqsubseteq_{n+1} X} \ (DefR)$$

  immediate

- (AndLi)

$$\frac{\theta_i \sqsubseteq_{n+1} X}{\theta_1 \sqcap \theta_2 \sqsubseteq_{n+1} X} \ (AndLi)$$

  (IH) yields $\theta_i \sqsubseteq_{n+1} \phi_X$, hence $\theta_1 \sqcap \theta_2 \sqsubseteq_{n+1} \phi_X$ can be derived using (AndLi)

- (DefL)

$$\frac{\phi_Y \sqsubseteq_{n+1} X}{Y \sqsubseteq_{n+1} X} \ (DefL)$$

  (IH) yields $\phi_Y \sqsubseteq_{n+1} \phi_X$, hence $Y \sqsubseteq_{n+1} \phi_X$ can be derived using (DefL)

- (Concept)

$$\frac{\theta \sqsubseteq_{n+1} \alpha \quad \beta \sqsubseteq_{n+1} X}{\theta \sqsubseteq_{n+1} X} \ (Concept) \text{ with } \alpha \sqsubseteq \beta \in \mathcal{F}$$

(IH) yields $\beta \sqsubseteq_{n+1} \phi_X$, hence $\theta \sqsubseteq_{n+1} \phi_X$ can be derived using (Concept).

The remaining part of the lemma is now easy to show. Indeed, $\theta \sqsubseteq_\infty \psi_1 \sqcap \psi_2$ iff $\theta \sqsubseteq_n \psi_1 \sqcap \psi_2$ for all $n \geq 0$ iff $\theta \sqsubseteq_n \psi_1$ and $\theta \sqsubseteq_n \psi_2$, for all $n \geq 0$ iff $\theta \sqsubseteq_\infty \psi_1$ and $\theta \sqsubseteq_\infty \psi_2$. Similarly, $\theta \sqsubseteq_\infty X$ iff $\theta \sqsubseteq_{n+1} X$ for all $n \geq -1$ iff $\theta \sqsubseteq_n \phi_X$ for all $n \geq 0$ iff $\theta \sqsubseteq_\infty \phi_X$. $\qquad\square$

Lemma 3.1.3. *Suppose that $n > 0$.*

$\theta \sqsubseteq_n \exists r.\psi$ *iff* *there exist α, β, ρ such that $\theta \sqsubseteq_n \alpha, \beta \Vdash \exists r.\rho$ and $\rho \sqsubseteq_n \psi$ for some subconcept ρ of either the T-Box, θ, or of $\exists r.\psi$, and α and β being such that either $\alpha = \beta = \theta$ or $\alpha \sqsubseteq \beta$ being a GCI from $\mathcal{F}$.*

Proof. For *if* direction one has to use (Ex) rule on $\rho \sqsubseteq_n \psi$ in order to obtain $\exists r.\rho \sqsubseteq_n \exists r.\psi$, then Lemma 3.1.1 to conclude $\beta \sqsubseteq_n \exists r.\psi$ and finally (Concept) rule to get the claim.

For the *only if* directions we proceed by induction on derivation $\theta \sqsubseteq_n \exists r.\psi$, where we distinguish cases depending on the last rule applied:

- (Ax) - base of induction $\exists r.\psi \sqsubseteq_n \exists r.\psi$ Here $\theta = \alpha = \beta$ and $\rho = \psi$
- (Ex)

$$\frac{\varphi \sqsubseteq_n \psi}{\exists r.\varphi \sqsubseteq_n \exists r.\psi} \ (Ex)$$

Immediate. Here $\theta = \alpha = \beta = \exists r.\varphi$ and $\rho = \varphi$
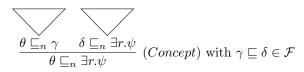- (AndL1)

$$\frac{\theta_1 \sqsubseteq_n \exists r.\psi}{\theta_1 \sqcap \theta_2 \sqsubseteq_n \exists r.\psi} \ (AndL1)$$

(IH) can yield $\theta_1 \Vdash \exists r.\rho,\ \rho \sqsubseteq_n \psi$, when the claim follows, since $\theta_1 \sqcap \theta_2 \Vdash \exists r.\rho$ can be derived using (AndL1). Alternatively, (IH) can yield $\theta_1 \sqsubseteq_n \alpha,\ \beta \Vdash \exists r.\rho,\ \rho \sqsubseteq_n \psi$, and this time claim follows again by (AndL1).
- (AndL2)

$$\frac{\theta_2 \sqsubseteq_n \exists r.\psi}{\theta_1 \sqcap \theta_2 \sqsubseteq_n \exists r.\psi} \ (AndL2)$$

(IH) can yield $\theta_2 \Vdash \exists r.\rho,\ \rho \sqsubseteq_n \psi$, when the claim follows, since $\theta_1 \sqcap \theta_2 \Vdash \exists r.\rho$ can be derived using (AndL2). Alternatively, (IH) can yield $\theta_2 \sqsubseteq_n \alpha,\ \beta \Vdash \exists r.\rho,\ \rho \sqsubseteq_n \psi$, and this time claim follows again by (AndLi).

26

- (DefL)

$$\frac{\phi_X \sqsubseteq_n \exists r.\psi}{X \sqsubseteq_n \exists r.\psi} \; (DefL)$$

(IH) can yield $\phi_X \Vdash \exists r.\rho$, $\rho \sqsubseteq_n \psi$, when the claim follows, since $X \Vdash \exists r.\rho$ can be derived using (DefL). Alternatively, (IH) can yield $\phi_X \sqsubseteq_n \alpha$, $\beta \Vdash \exists r.\rho$, $\rho \sqsubseteq_n \psi$. Again, claim follows applying (DefL).

- (Concept)

$$\frac{\theta \sqsubseteq_n \gamma \quad \delta \sqsubseteq_n \exists r.\psi}{\theta \sqsubseteq_n \exists r.\psi} \; (Concept) \text{ with } \gamma \sqsubseteq \delta \in \mathcal{F}$$

(IH) yields $\delta \sqsubseteq_n \alpha$, $\beta \Vdash \exists r.\rho$, $\rho \sqsubseteq_n \psi$, where δ, α and β may be identical. Now, the claim follows since $\theta \sqsubseteq_n \alpha$ can be derived using (Concept) on $\theta \sqsubseteq_n \gamma$ and $\delta \sqsubseteq_n \alpha$.

All the other cases are not applicable. $\qquad\qquad\square$

A simple corollary of the previous lemma is the following one.

Lemma 3.1.4.

$\theta \sqsubseteq_\infty \exists r.\psi \quad$ *iff* $\quad$ *there exist α, β, ρ such that $\theta \sqsubseteq_\infty \alpha, \beta \Vdash \exists r.\rho$ and $\rho \sqsubseteq_\infty \psi$ for some subconcept ρ of either the T-Box, θ, or of $\exists r.\psi$, and α and β being such that either $\alpha = \beta = \theta$ or $\alpha \sqsubseteq \beta$ being a GCI from $\mathcal{F}$.*

Proof. Having $\theta \sqsubseteq_\infty \exists r.\psi$ means that $\theta \sqsubseteq_n \exists r.\psi$ holds for every n. Previous lemma furnishes α, β and ρ such that $\theta \sqsubseteq_n \alpha, \beta \Vdash \exists r.\rho$ and $\rho \sqsubseteq_n \psi$ for every such n. However, there is only a finite choice for α, β and ρ (they are subconcepts occurring in the TBox), thus there is a combination of α, β and ρ for which $\theta \sqsubseteq_n \alpha, \beta \Vdash \exists r.\rho$ and $\rho \sqsubseteq_n \psi$ holds for infinitely many n. By the definition of $\sqsubseteq_\infty$ and the fact that $\sqsubseteq_{n+1} \subseteq \sqsubseteq_n$, for those α, β and ρ it will hold $\theta \sqsubseteq_\infty \alpha, \beta \Vdash \exists r.\rho$ and $\rho \sqsubseteq_\infty \psi$. $\qquad\square$

3.2 Soundness

It was announced before that provability of the relation $\sqsubseteq_\infty$ will decide subsumption with respect to a hybrid TBox. In order to prove that, soundness and completeness of HC have to be shown. In this context, soundness means that if relation $\phi \sqsubseteq_\infty \psi$ is provable, then $\phi \sqsubseteq \psi$ has to be valid in greatest fixpoint

semantics that is defined for a given hybrid TBox. Completeness means that for every valid subsumption $\phi \sqsubseteq \psi$ modulo appropriate semantics there has to be a proof of $\phi \sqsubseteq_\infty \psi$ in the given calculus.

The standard idea for showing soundness of a calculus by showing validity of the axioms and soundness of the rules fails in this case. The reason is simple. Namely, the rule (Start) is not valid! The 'problematic' rule is the (DefR) rule that connects derivations of different 'generations' of $\sqsubseteq_n$ relation, $\sqsubseteq_n$ and $\sqsubseteq_{n+1}$, and which can lead a bottom-up derivation to the undesirable (Start) axiom. Hence, another approach has to be used.

Usually, when analyzing calculi, one is interested in CUT-elimination. CUT rule in HC calculus actually is the claim about the transitivity of $\sqsubseteq_n$, i.e. whenever there is a proof of $\phi \sqsubseteq_\infty \psi$ and $\psi \sqsubseteq_\infty \theta$, there will be a proof of $\phi \sqsubseteq_\infty \theta$. CUT-elimination is a strong claim, and a crucial feature of many calculi in standard proof theory that provides subformula property, analyticity of a calculus, completeness, interpolation theorems, normalization of the proofs, etc. However, the fact that only subconcepts occurring in the TBox can occur in derivations in our case makes most of those features unnecessary. By this, we mean that when performing bottom-up construction of a proof, there is always finite choice of premises in each step of the construction. This is not the case in bottom up proof construction in calculi for classical logic with CUT, for instance. In the case of HC, the subformula property of the calculus (concepts in the premises are subconcepts of the concepts in the conclusion) is already lost with the rule (Concept) that introduces new concept descriptions. Notice also that adding the CUT rule into HC system would not change the set of provable relations. We will show completeness of the HC calculus, so whenever there is a proof of $\phi \sqsubseteq_\infty \psi$ and $\psi \sqsubseteq_\infty \theta$, there will be a proof of $\phi \sqsubseteq_\infty \theta$, due to the transitivity of set-inclusion.

The following lemma shows soundness of $\Vdash$ relation.

Lemma 3.2.1. *Let $\mathcal{I}$ be a greatest fixpoint model of a hybrid TBox $(\mathcal{F}, \mathcal{T})$. If $\phi \Vdash \psi$ then $\mathcal{I}(\phi) \subseteq \mathcal{I}(\psi)$.*

Proof. Again, proof is done by induction on the derivation of $\phi \Vdash \psi$. We distinguish cases depending on the last rule applied.

1. (Ax) - base of induction: $\phi \Vdash \phi$ implies $\mathcal{I}(\phi) \subseteq \mathcal{I}(\phi)$

2. (AndLi) (for $i = 1, 2$):

$$\frac{\theta_i \sqsubseteq_n \psi}{\phi_1 \sqcap \phi_2 \sqsubseteq_n \psi} \; (AndLi)$$

induction hypothesis yields $\mathcal{I}(\phi_i) \subseteq \mathcal{I}(\psi)$, and since $\mathcal{I}$ is a model, one has $\mathcal{I}(\phi_1 \sqcap \phi_2) \subseteq \mathcal{I}(\phi_i) \subseteq \mathcal{I}(\psi)$

28

3. (DefL):

$$\frac{\phi_X \sqsubseteq_n \psi}{X \sqsubseteq_n \psi} \ (DefL)$$

the induction hypothesis yields $\mathcal{I}(\phi_X) \subseteq \mathcal{I}(\psi)$, and since $\mathcal{I}$ is a model of $\mathcal{T}$, and $X \equiv \phi_X \in \mathcal{T}$, one has $\mathcal{I}(X) = \mathcal{I}(\phi_X) \subseteq \mathcal{I}(\psi)$

4. (Concept):

$$\frac{\phi \sqsubseteq_n \alpha \qquad \beta \sqsubseteq_n \psi}{\phi \sqsubseteq_n \psi} \ (Concept) \text{ with } \alpha \sqsubseteq \beta \in \mathcal{F}$$

induction hypothesis yields $\mathcal{I}(\phi) \subseteq \mathcal{I}(\alpha), \mathcal{I}(\beta) \subseteq \mathcal{I}(\psi)$, and since $\mathcal{I}$ is a model of $\mathcal{F}$, and $\alpha \sqsubseteq \beta \in \mathcal{F}$, one has $\mathcal{I}(\phi) \subseteq \mathcal{I}(\alpha) \subseteq \mathcal{I}(\beta) \subseteq \mathcal{I}(\psi)$.

$\square$

So far, only the relation $\Vdash$ is known to be sound in the sense of the previous lemma. The following two lemmas extends this property to important subsets of the relations $\sqsubseteq_n$, namely to those pairs where second component is a concept description that does not contain defined concepts, i.e. it has no subconcept that is element of N_{def}. Notice that in this calculus it does not hold that $\theta \sqsubseteq_n P$ iff $\theta \Vdash P$, with $P \in N_{prim}$. As an example, consider the TBox $(\mathcal{F}, \mathcal{T})$, $\mathcal{F} = \{\exists r.P \sqsubseteq Q\}$, $\mathcal{T} = \{X \equiv P\}$. Then, there is single derivation of $\exists r.X \sqsubseteq_n Q$:

$$\cfrac{\cfrac{\cfrac{\cfrac{\overline{P \sqsubseteq_n P}\ (Ax)}{X \sqsubseteq_n P}\ (DefL)}{\exists r.X \sqsubseteq_n \exists r.P}\ (Ex) \qquad \overline{Q \sqsubseteq_n Q}\ (Ax)}{\exists r.X \sqsubseteq_n Q}\ (Concept)}{}$$

However, it is obvious that $\exists r.X \not\Vdash Q$, due to the usage of the rule (Ex).

Lemma 3.2.2. *Suppose $n > 0$. Let α be a concept description formed using only concepts from N_{prim}, i.e. α is a concept description that does not have any subconcept from N_{def}. Then the proof tree for $\theta \sqsubseteq_n \alpha$ will only include expressions of the form $\psi \sqsubseteq_n \beta$, with β being a concept description over the concepts from N_{prim} only.*

Proof. Proof is done by induction on derivation of $\theta \sqsubseteq_n \alpha$. We distinguish cases depending on the last rule applied, which can be one of the (Ax), (Top), (AndL1), (AndL2), (DefL) and (Concept). Base of the induction are cases (Ax) and (Top).

- (Ax) $\alpha \sqsubseteq_n \alpha$. There is nothing to prove, since obviously proof for $\theta \sqsubseteq_n \alpha$ contains no sequents with right-hand sides having subconcepts from N_{def}.

- (Top) $\theta \sqsubseteq_n \top$. Again, there is nothing to prove, since obviously proof for $\theta \sqsubseteq_n \alpha$ contains no sequents with right-hand sides having subconcepts from N_{def}.

- (AndLi) (for i=1, 2)

$$\frac{\overset{\displaystyle\mathcal{P}}{\theta_i \sqsubseteq_n \exists r.\psi}}{\theta_1 \sqcap \theta_2 \sqsubseteq_n \alpha} \; (AndLi)$$

(IH) can be applied to the premise of the (AndLi) rule to conclude that $\mathcal{P}$, and therefore entire proof of $\theta \sqsubseteq_n \alpha$ contains no sequents with right-hand sides having subconcepts from N_{def}.

- (DefL)

$$\frac{\overset{\displaystyle\mathcal{P}}{\phi_X \sqsubseteq_n \alpha}}{X \sqsubseteq_n \alpha} \; (DefL)$$

(IH) can be applied to the premise of the (DefL) rule to conclude that $\mathcal{P}$, and therefore entire proof of $\theta \sqsubseteq_n \alpha$ contains no sequents with right-hand sides having subconcepts from N_{def}.

- (Concept)

$$\frac{\overset{\displaystyle\mathcal{P}'}{\theta \sqsubseteq_n \gamma} \quad \overset{\displaystyle\mathcal{P}''}{\delta \sqsubseteq_n \alpha}}{\theta \sqsubseteq_n \exists r.\psi} \; (Concept) \text{ with } \gamma \sqsubseteq \delta \in \mathcal{F}$$

γ is a subconcept occuring in $\mathcal{F}$, thus it contains no subconcepts from N_{def}. Hence, (IH) can be applied to both of the premises of the (Concept) rule to conclude that $\mathcal{P}'$, $\mathcal{P}''$, and therefore entire proof of $\theta \sqsubseteq_n \alpha$ contains no sequents with right-hand sides having subconcepts from N_{def}.

$\square$

Lemma 3.2.3. *Suppose $n > 0$. Let $\mathcal{I}$ be a greatest fixpoint model of a hybrid TBox $(\mathcal{F}, \mathcal{T})$ and α be a concept description formed using only concepts from N_{prim}, i.e. α is a concept description that does not contain any occurrence of a concept name from N_{def}. If $\theta \sqsubseteq_n \alpha$, then $I(\theta) \subseteq I(\alpha)$.*

Proof. By Lemma 3.2.2 proof of $\theta \sqsubseteq_n \alpha$ contains no sequents with right-hand sides having subconcepts from N_{def}. In particular, the proof contains no application of the rule (DefR). This means that the proof for $\theta \sqsubseteq_n \alpha$ has only instances of (Ax) and (Top) as leafs. Indeed, absence of the (DefR) rule implies the conclusion that all the rule applications from the proof have the same relation $\sqsubseteq_n$ in conclusion as well as in the premises, i.e. it is impossible to regress to $\sqsubseteq_m$, with $m \le n$, in bottom-up construction of the proof tree. Thus, so every branch of the proof tree for $\theta \sqsubseteq_n \alpha$ has to end with a sequent with the same $\sqsubseteq_n$ relation.

Since all rules of the calculus except (Start) are sound in the usual sense, the same approach to proving soundness as the one used in Lemma 3.2.1 can be applied in this case. In other words, given a model $\mathcal{I}$ of the hybrid TBox, simple inductive argument proves that for every $\rho \sqsubseteq_n \xi$ in the proof tree that has only (Ax) and (Top) instances as leaves, $\mathcal{I}(\rho) \subseteq \mathcal{I}(\xi)$. In particular, $I(\theta) \subseteq I(\alpha)$. □

Finally, previous lemmas allow for a proof of soundness to go through.

Theorem 3.2.1. *(Soundness) If $\theta \sqsubseteq_\infty \phi$ then $\mathcal{I}(\theta) \subseteq \mathcal{I}(\phi)$ for any greatest fixpoint model $\mathcal{I}$ of a hybrid $\mathcal{EL}$-TBox $(\mathcal{F}, \mathcal{T})$.*

Proof. We want to show that for all subconcepts θ, ϕ of the hybrid $\mathcal{EL}$-TBox $(\mathcal{F}, \mathcal{T})$ and any model $\mathcal{I}$ of $(\mathcal{F}, \mathcal{T})$ we have

$$\theta \sqsubseteq_\infty \phi \Longrightarrow \mathcal{I}(\theta) \subseteq \mathcal{I}(\phi)$$

Equivalently, we may try to prove that for each ϕ and interpretation $\mathcal{I}$ one has

$$(\bigcup_{\theta \sqsubseteq_\infty \phi} \mathcal{I}(\theta)) \subseteq \mathcal{I}(\phi)$$

Writing $\mathcal{J}(\phi) := \bigcup_{\theta \sqsubseteq_\infty \phi} \mathcal{I}(\theta)$ for the left hand side this would follow if we can prove that $\mathcal{J}$ satisfies (dis)equations (1)-(5) form the Definition 2.2.6 of greatest fixpoint semantics. We restrict here $\mathcal{J}$ to have only subconcepts occurring in the TBox as a domain. This is because by its definition, $\mathcal{J}$ would interpret all non-subconcepts as $\varnothing$, and this may contradict the conditions of the Definition 2.2.6. However, due to the Proposition 2.2.3, this restriction of $\mathcal{J}$ to subformulae does not impose a problem since we can restrict our attention to the subconcepts occurring in the TBox.

Proofs of $\mathcal{J}$ fulfilling (1)-(5) from the Definition 2.2.6 are given below.

- (1): Since $\top \sqsubseteq_\infty \top$ by the rule (Ax), $\mathcal{J}(\top) = \mathcal{I}(\top)$
- (2): Since $P \sqsubseteq_\infty P$ by the rule (Ax), $\mathcal{J}(P) \supseteq \mathcal{I}(P)$. Lemma 3.2.3, on the other hand, states that, for any ϕ, $\phi \sqsubseteq_\infty P$ implies $\mathcal{I}(\phi) \subseteq \mathcal{I}(P)$, hence $\mathcal{J}(P)$ is union of subsets of $\mathcal{I}(P)$, therefore $\mathcal{J}(P) \subseteq \mathcal{I}(P)$. Combined with the opposite direction, one has $\mathcal{J}(P) = \mathcal{I}(P)$.

31

- (3): This disequation follows immediately from Lemma 3.1.2. Indeed, if $x \in \mathcal{J}(X)$, then $x \in \mathcal{I}(\theta)$, for some θ s.t. $\theta \sqsubseteq_\infty X$. But then, by Lemma 3.1.2, $\theta \sqsubseteq_\infty \phi_X$, therefore $x \in \mathcal{J}(\phi_X)$, i.e., $\mathcal{J}(X) \subseteq \mathcal{J}(\phi_X)$

- (4) This equation follows immediately from Lemma 3.1.2. For some x of the domain, it follows $x \in \mathcal{J}(\phi \sqcap \psi)$ iff $x \in \mathcal{I}(\theta)$, with $\theta \sqsubseteq_\infty \phi \sqcap \psi$. According to the lemma, the last is equivalent to $x \in \mathcal{I}(\theta)$, with $\theta \sqsubseteq_\infty \phi$ and $\theta \sqsubseteq_\infty \psi$, which is equivalent to $x \in \mathcal{J}(\phi)$ and $x \in \mathcal{J}(\psi)$, or $x \in \mathcal{J}(\phi) \cap \mathcal{J}(\psi)$

- (5) If $x \in \mathcal{J}(\exists r.\phi)$ then by definition of $\mathcal{J}$ there exists a formula θ such that $\theta \sqsubseteq_\infty \exists r.\phi$ and $x \in \mathcal{I}(\theta)$. By Lemma 3.1.4, this means that we can find α, β and ρ such that $\theta \sqsubseteq_\infty \alpha, \beta \Vdash \exists r.\rho$ and $\rho \sqsubseteq_\infty \phi$, with $\alpha \sqsubseteq \beta$ being a GCI of $\mathcal{F}$ or $\alpha = \beta = \theta$. Concept α is a subconcept occurring in $\mathcal{F}$, thus Lemma 3.2.3 can be applied to obtain $x \in \mathcal{I}(\alpha)$, and therefore, $x \in \mathcal{I}(\beta)$. Lemma 3.2.1 yields $x \in \mathcal{I}(\exists r.\rho)$ so there exists $y \in \mathcal{I}(\rho)$ with $\mathcal{I}(r)(x, y)$. Since $\rho \sqsubseteq_\infty \phi$, y will belong to $\mathcal{J}(\phi)$, therefore, if $x \in \mathcal{J}(\exists r.\phi)$ then $x \in \{z \mid \exists y \in \mathcal{J}(\phi).\mathcal{I}(r)(z, y)\}$.

  Conversely, if $x \in \{z \mid \exists y \in \mathcal{J}(\phi).\mathcal{I}(r)(z, y)\}$, then there is $y \in \mathcal{J}(\phi)$ such that $\mathcal{I}(r)(x, y)$. This means that there is a concept θ such that $y \in \mathcal{I}(\theta)$, $\theta \sqsubseteq_\infty \phi$ and $\mathcal{I}(r)(x, y)$. It is clear that $\theta \sqsubseteq_\infty \phi$ implies $\exists r.\theta \sqsubseteq_\infty \exists r.\phi$ (for every n, derivation of $\theta \sqsubseteq_\infty \phi$ can be extended to a derivation of $\exists r.\theta \sqsubseteq_\infty \exists r.\phi$ by applying (Ex)), thus the last statement implies $x \in \mathcal{I}(\exists r.\theta)$ and $\exists r.\theta \sqsubseteq_\infty \exists r.\phi$. This is exactly the claim $x \in \mathcal{J}(\exists r.\phi)$.

$\square$

It is now obvious why it was important to characterize the proofs in the way it was done in Lemma 3.1.2 and Lemma 3.1.3. Probably the most problematic part of the proof was handling the existential restriction case, and characteristics of the relation $\Vdash$ came into play here, as well as the feature of derivations with concepts over N_{prim} on the right-hand side, as described in Lemma 3.2.3.

3.3 Completeness

The proof that will be presented uses most of the properties shown before. The idea is to consider an interpretation that uses subconcepts occurring in the TBox (with the subsumption query added) as the domain and interprets a concept description by the set of provably subsumed subconcepts. The fact that this is well defined interpretation will yield completeness proof.

Theorem 3.3.1. *(Completeness) If $A \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}} B$ then $A \sqsubseteq_\infty B$.*

Proof. Define:

$$\mathcal{D}_\mathcal{I} \;=\; \text{subconcepts occurring in the TBox}$$

$$\mathcal{I}(r)(\phi,\psi) \;\Leftrightarrow\; \phi \sqsubseteq_\infty \alpha \text{ and } \beta \Vdash \exists r.\psi \text{ for some } \alpha \sqsubseteq \beta \text{ being a GCI in } \mathcal{F},$$
$$\text{or } \alpha = \beta = \phi$$

$$\mathcal{I}(\phi) \;=\; \begin{cases} \{\psi | \psi \sqsubseteq_\infty \phi\}, & \text{if } \phi \text{ is a subconcept} \\ & \text{occurring in } (\mathcal{F},\mathcal{T}) \\ \mathcal{I}(\phi_1) \cap \mathcal{I}(\phi_2), & \text{if } \phi \text{ does not occur in } (\mathcal{F},\mathcal{T}) \\ & \text{and } \phi = \psi_1 \sqcap \psi_2 \\ \{\psi | \text{ there is } \theta \in \mathcal{I}(\phi_1) & \text{if } \phi \text{ does not occur in } (\mathcal{F},\mathcal{T}) \\ \text{such that } \mathcal{I}(r)(\psi,\theta)\}, & \text{and } \phi = \exists r.\psi_1 \end{cases}$$

We proceed to show that $\mathcal{I}$ is a gfp-model of the hybrid TBox $(\mathcal{F},\mathcal{T})$.

In order to show that $\mathcal{I}$ is a gfp model of $(\mathcal{F},\mathcal{T})$ we need to prove that $\mathcal{I}$ is a well defined interpretation, that it is a descriptive model of $\mathcal{T}$ and $\mathcal{F}$, and that it is the gfp model of $\mathcal{T}$.

In order to show that $\mathcal{I}$ is an interpretation, it has to be shown that indeed:

- $\mathcal{I}(\phi_1 \sqcap \phi_2) = \mathcal{I}(\phi_1) \cap \mathcal{I}(\phi_2)$: If $\phi_1 \sqcap \phi_2$ is not a subconcept occurring in the Tbox, claim follows by definition of $\mathcal{I}$. Assume now that $\phi_1 \sqcap \phi_2$ occurs in $(\mathcal{F},\mathcal{T})$. For some θ of the domain, it holds that $\theta \in \mathcal{I}(\phi \sqcap \psi)$ iff $\theta \sqsubseteq_\infty \phi \sqcap \psi$. From Lemma 3.1.2, the last is equivalent to $\theta \sqsubseteq_\infty \phi$ and $\theta \sqsubseteq_\infty \psi$, which is equivalent to $\theta \in \mathcal{I}(\phi)$ and $\theta \in \mathcal{I}(\psi)$, or $\theta \in \mathcal{I}(\phi) \cap \mathcal{I}(\psi)$

- $\mathcal{I}(\exists r.\phi_1) = \{\varphi \mid \exists \xi \in \mathcal{I}(\phi_1) : (\varphi,\xi) \in \mathcal{I}(r)\}$: If $\exists r.\phi_1$ is not a subconcept occurring in the Tbox, claim follows by definition of $\mathcal{I}$. Assume now that $\exists r.\phi_1$ occurs in $(\mathcal{F},\mathcal{T})$. If $\psi \in \mathcal{I}(\exists r.\phi_1)$ then $\psi \sqsubseteq_\infty \exists r.\phi_1$, so by Lemma 3.1.3 there exist α,β and ρ such that $\psi \sqsubseteq_n \alpha, \beta \Vdash \exists r.\rho$ and $\rho \sqsubseteq_n \phi_1$, with $\alpha \sqsubseteq \beta$ being a GCI of $\mathcal{F}$ or $\alpha = \beta = \psi$. Again, as in the proof of Lemma 3.1.3 may furnish a different α,β and ρ for each n, for a global one it suffices to take one (of the finitely many possible) that occurs for infinitely many n. This means that $\mathcal{I}(r)(\psi,\rho)$ and $\rho \in \mathcal{I}(\phi_1)$ as required.

  Conversely, if $\mathcal{I}(r)(\psi,\rho)$ and $\rho \sqsubseteq_\infty \phi_1$ for some ψ and ρ, then there exist α,β and ρ such that $\psi \sqsubseteq_n \alpha, \beta \Vdash \exists r.\rho$ $\rho \sqsubseteq_n \phi_1$, with $\alpha \sqsubseteq \beta$ being a GCI of $\mathcal{F}$ or $\alpha = \beta = \psi$, for every n. If $\alpha = \beta = \psi$, we get $\psi \Vdash \exists r.\rho$, and then $\psi \sqsubseteq_\infty \exists r.\phi_1$ by (Ex) and Lemma 3.1.1. On the other hand, if $\alpha \sqsubseteq \beta$ is a GCI of $\mathcal{F}$, then $\beta \Vdash \exists r.\phi_1$ by (Ex) and Lemma 3.1.1, and then $\psi \sqsubseteq_\infty \exists r.\phi_1$ by (Concept).

In order to show that $\mathcal{I}$ is a descriptive model of $\mathcal{T}$, we have to show that $\mathcal{I}(X) \equiv \mathcal{I}(\phi_X)$ for every definition $X \equiv \phi_X$ of $\mathcal{T}$.

If $\theta \in \mathcal{I}(X)$, then $\theta \sqsubseteq_\infty X$. But then, by Lemma 3.1.2, $\theta \sqsubseteq_\infty \phi_X$, therefore $\theta \in \mathcal{I}(\phi_X)$, i.e., $\mathcal{I}(X) \subseteq \mathcal{I}(\phi_X)$. Converse is also true, i.e., if $\theta \in \mathcal{I}(\phi_X)$, then $\theta \sqsubseteq_\infty \phi_X$ and Lemma 3.1.2 yields $\theta \sqsubseteq_\infty X$. Hence, $\theta \in \mathcal{I}(X)$ and $\mathcal{I}(\phi_X) \subseteq \mathcal{I}(X)$.

Further more, the interpretation $\mathcal{I}$ satisfies $\mathcal{F}$. Indeed, if $\psi \sqsubseteq \phi \in \mathcal{F}$, and $\alpha \in \mathcal{I}(\psi)$, one has $\alpha \sqsubseteq_\infty \psi$ and $\phi \sqsubseteq_\infty \phi$, which gives $\alpha \sqsubseteq_\infty \phi$ using (Concept). Therefore $\alpha \in \mathcal{I}(\phi)$, i.e $\psi \sqsubseteq \phi \in \mathcal{F}$ implies $\mathcal{I}(\psi) \subseteq \mathcal{I}(\phi)$.

It remains to show that $\mathcal{I}$ is indeed the greatest interpretation. To show this, assume that $\mathcal{J}$ is a function from subconcepts occurring in $(\mathcal{F}, \mathcal{T})$ to subsets of $\mathcal{D}_\mathcal{I}$ that also satisfies (1)-(5) and coincides with $\mathcal{I}$ on elements of N_{def}. By Proposition 2.2.3, it suffice to show that $\mathcal{J}(\psi) \subseteq \mathcal{I}(\psi)$ for all subconcepts ψ occurring in the TBox. In other words, if $\phi \in \mathcal{J}(\psi)$ then $\phi \sqsubseteq_n \psi$ for all $n \in \mathbb{N}$. We show this by induction on n and a subsidiary induction on ψ.

- If $\psi = P$, $\mathcal{J}(P) = \mathcal{I}(P)$ by assumption on $\mathcal{J}$.
- If $\psi = \top$ then $\mathcal{J}(P) = \mathcal{I}(\top) = \mathcal{D}_\mathcal{I}$
- If $\psi = \psi_1 \sqcap \psi_2$ then $\phi \in \mathcal{J}(\psi_i)$ and we may inductively assume that $\phi \sqsubseteq_n \psi_i$ so $\phi \sqsubseteq_n \psi$ by (AndR).

- If $\psi = \exists r.\psi_1$ then there exists a formula $\rho \in \mathcal{J}(\psi_1)$ and α, β such that $\phi \sqsubseteq_\infty \alpha$, $\beta \Vdash \exists r.\rho$, with $\alpha \sqsubseteq \beta \in \mathcal{F}$ or $\alpha = \beta = \phi$. The subordinate induction hypothesis applied to ψ_1 yields $\rho \sqsubseteq_n \psi_1$ and thus $\exists r.\rho \sqsubseteq_n \exists r.\psi_1$ by rule (Ex) and thus $\beta \sqsubseteq_n \psi$ by Lemma 3.1.1. One more application of (Concept) gives $\phi \sqsubseteq_n \psi$
- If, finally, $\psi = X$ we distinguish two cases. If $n = 0$ then $\phi \sqsubseteq_n \psi$ by (Start). Otherwise, $n = n' + 1$ and $\phi \in \mathcal{J}(\phi_X)$ since $\mathcal{J}(X) \subseteq \mathcal{J}(\phi_X)$. The outer induction hypothesis yields $\phi \sqsubseteq_{n'} \phi_X$ and we obtain $\phi \sqsubseteq_n X$ by (DefR).

Finally, if $A \sqsubseteq_{\text{gfp}, \mathcal{F}, \mathcal{T}} B$, since $\mathcal{I}$ is indeed a greatest fixpoint model for the hybrid TBox, it has to be $\mathcal{I}(A) \subseteq \mathcal{I}(B)$, i.e. $\{\zeta \mid \zeta \sqsubseteq_\infty A\} \subseteq \{\eta \mid \eta \sqsubseteq_\infty B\}$. In particular, $A \in \{\zeta \mid \zeta \sqsubseteq_\infty A\}$, hence $A \in \{\eta \mid \eta \sqsubseteq_\infty B\}$, which gives $A \sqsubseteq_\infty B$. $\qquad\square$

3.4 Decision procedure

Previous sections provided tools for developing a decision procedure for the subsumption problem. The decision procedure will compute the relation $\sqsubseteq_\infty$, i.e. it does not only decide a particular subsumption relationship, it also provides all pairs ϕ, ψ such that $\phi \sqsubseteq_\infty \psi$. An important fact that enables polynomial runtime is that we are able to restrict our attention to the subconcepts occurring in the TBox. Consideration of all possible subsumption relationships between concepts that can be built form the subconcepts of the TBox is, of course, exponential, since one has to deal (among other things) with the power set of the set of all subconcepts.

Theorem 3.4.1. *The relation $\sqsubseteq_{gfp,\mathcal{F},\mathcal{T}}$ (subsumption under the hybrid semantics) is decidable in polynomial time.*

Proof. Completeness of HC reduces deciding of $\sqsubseteq_{\text{gfp},\mathcal{F},\mathcal{T}}$ relation to deciding of $\sqsubseteq_\infty$. As already mentioned, one has $\sqsubseteq_{n+1}\subseteq\sqsubseteq_n$, so one can compute these relations by iteration. We maintain two tables of polynomial size to hold the relations $\sqsubseteq_n$ and $\sqsubseteq_{n-1}$. Size of these tables is quadratic to the number of subconcepts occurring in the hybrid TBox, thus polynomial (quadratic in the size of the TBox). Initially, we put $n = 0$ and set $\sqsubseteq_n$ to be the total relation in view of rule (Start). If relation $\sqsubseteq_{n-1}$ has already been computed we compute $\sqsubseteq_n$ from it by iteration, deciding (in polynomial time) all of the $\phi \sqsubseteq_n \psi$, and so forth. For every n, $\sqsubseteq_n$ can be computed by bottom-up iteration or dynamic programming in the following way: one sets up an array $A_n[\phi, \psi]$ that contains a boolean entry for every pair of subconcepts ϕ and ψ. Initially all entries are set to 0. For every instance of the proof rule with premises $\sigma_i \sqsubseteq_n \tau_i$, $(0 \leq i < 3)$ and the conclusion $\phi \sqsubseteq_n \psi$, with $A_n[\sigma_i, \tau_i] = 1$, set $A_n[\phi, \psi]$ to 1. Also, if $\phi \sqsubseteq_{n-1} \phi_X$, set $A[\phi, X]$ to 1. First non-zero values in A_n will be set form the instances of (Ax), (Top) and (DefR). One stops when the procedure stabilizes. Notice that for computing A_n one needs to know only the values form A_{n-1}, therefore, as mentioned, it suffice to store only two arrays for n and $n - 1$. Number of operations needed for computing $\sqsubseteq_n$ (stabilized A_n) is polynomial in the size of A_n. Indeed, the number of different instances of rules in HC is polynomial in the size of the TBox, since every rule contains at most 4 subconcepts occurring in the TBox, and total number of subconcepts is linear in the size of the TBox. Further more, testing for premises of a rule instance can be done in cycles, and the number of these cycles is bounded by the size of A_n, thus quadratic to the size of the input. Algorithms for computing $\sqsubseteq_n$ are again discussed in Chapter 5.

After computing $\sqsubseteq_n$ (A_n), we proceed with $\sqsubseteq_{n+1}$. As soon as we have reached an n_0 for which $\sqsubseteq_{n_0+1}=\sqsubseteq_{n_0}$ we may stop since for such n_0 one has $\sqsubseteq_\infty=\sqsubseteq_{n_0}$. But if $\sqsubseteq_{n+1}\subsetneq\sqsubseteq_n$ then $\sqsubseteq_{n+1}$ must contain at least one pair less than $\sqsubseteq_n$ which implies that n_0 exists and is polynomial in $|\mathcal{T}| + |\mathcal{F}|$. The overall time complexity of the procedure is thus bounded by . This implies a polynomial runtime of the decision procedure. $\qquad\square$

Chapter 4

Proof-theoretic computation of least common subsumer in $\mathcal{EL}$ w.r.t. hybrid TBoxes

This chapter describes a proof-theoretic technique to calculate the least-common subsumer of two defined concepts with respect to hybrid TBoxes. Notion of least common subsumer w.r.t. hybrid $\mathcal{EL}$ TBoxes is introduced and a method for computing it is presented in Section 4.1. The correctness of the computation is shown in Section 4.2. Through entire chapter we use proof-theoretic techniques developed in the previous chapters to show many of the properties of the algorithm, and its correctness. However, soundness and completeness of the devised calculus HC enables us to use reasoning on the semantics if necessary, and semantical proofs will be given it will in several occasions.

4.1 Computing least common subsumer in $\mathcal{EL}$ w.r.t. hybrid TBoxes

We start by giving formal definition of the notion of an extension/restriction of an interpretation, that will be used in this chapter.

Definition 4.1.1. Given an interpretation $\mathcal{I} = (\mathcal{D}_{\mathcal{I}}, \mathcal{I}(\cdot))$ defined on the set of primitive concepts N_{prim}, the set of role names N_{role} and a set of defined concepts N_{def}, we say that an interpretation $\mathcal{J} = (\mathcal{D}_{\mathcal{I}}, \mathcal{J}(\cdot))$ defined on the set of primitive concepts N'_{prim}, the set of role names N'_{role} and a set of defined concepts N'_{def}, is an *extension* of $\mathcal{I}$ if $N_{prim} \subseteq N'_{prim}$, $N_{def} \subseteq N'_{def}$ and $N_{role} \subseteq N'_{role}$ and $\mathcal{J}(x) = \mathcal{I}(x)$ for all x in N_{prim}, N_{role} and N_{def}. In this case we say that $\mathcal{I}$ is a *restriction* of $\mathcal{J}$.

Notion of extension of an interpretation given here is general, and we will, if not stated else, consider the extensions/restrictions where $N_{role} = N'_{role}$.

We proceed by introducing the notion of a conservative extension of a hybrid $\mathcal{EL}$ TBox.

Definition 4.1.2. 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.

One important property of conservative extensions of hybrid TBoxes is given in the following lemma:

Lemma 4.1.1. *Let $(\mathcal{F}, \mathcal{T})$ be a hybrid TBox, and $(\mathcal{F}, \mathcal{T}')$ its arbitrary conservative extension. Then $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T})} \psi$ iff $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}')} \psi$, for all subconcepts ϕ, ψ occurring in $(\mathcal{F}, \mathcal{T})$.*

Proof. Assume $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T})} \psi$. Every derivation of $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T})} \psi$ is also a derivation of $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}')} \psi$ (modulo renaming of the superscripts), and thus $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}')} \psi$.

Assume now $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}')} \psi$. Then, every derivation of $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}')} \psi$ does not involve concepts from $N_{def}^{\mathcal{T}'} \setminus N_{def}^{\mathcal{T}}$, therefore it can be transformed to a derivation of $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T})} \psi$ merely by changing the superscripts from $(\mathcal{F}, \mathcal{T}')$ to $(\mathcal{F}, \mathcal{T})$. Indeed, looking at the derivations in a bottom-up fashion, the only rules that may introduce concepts that do not occur in conclusions are (DefR), (DefL) and (Concept). However, in all three cases, the only concepts that are introduced are those over N_{prim}, N_{role} and $N_{def}^{\mathcal{T}}$, i.e., concepts form $N_{def}^{\mathcal{T}'} \setminus N_{def}^{\mathcal{T}}$ cannot occur in the premises. Now, $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T})} \psi$ follows from $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T})} \psi$ for every n. □

Previous lemma has an important corollary that for a hybrid TBox $(\widehat{\mathcal{F}}, \widehat{\mathcal{T}})$, its conservative extension $(\widehat{\mathcal{F}}, \widehat{\mathcal{T}}')$, and concepts ϕ and ψ form $(\widehat{\mathcal{F}}, \widehat{\mathcal{T}})$, it holds $(\widehat{\mathcal{F}}, \widehat{\mathcal{T}}) \models \phi \sqsubseteq \psi$ iff $(\widehat{\mathcal{F}}, \widehat{\mathcal{T}}') \models \phi \sqsubseteq \psi$. This is due to the fact that $(\hat{\mathcal{F}}, \hat{\mathcal{T}}) \models \phi \sqsubseteq \psi$ iff $\phi \sqsubseteq_\infty^{(\hat{\mathcal{F}}, \hat{\mathcal{T}})} \psi$ and the previous lemma. Notice that the same conclusion could be obtained by semantic argumentation from the definition of conservative extensions. Still, we chose to employ HC calculus instead in order to demonstrate its usage.

The following definition introduces notion of least-common subsumer in the hybrid setting.

Definition 4.1.3. (Hybrid lcs) Let $(\mathcal{F}, \mathcal{T}_1)$ be a hybrid $\mathcal{EL}$ TBox and $X, Y \in N_{def}^{\mathcal{T}_1}$. 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 X, Y in $(\mathcal{F}, \mathcal{T}_1)$ iff the following conditions hold:

1. $X \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_2} Z$ and $Y \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_2} Z$; 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 $X \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_3} D$ and $Y \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_3} D$ then $Z \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_3} D$.

Notice that 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 suffice to consider only newly defined concepts for testing the condition 2. of the definition above. Indeed, if we want to test wether $Z \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_3} D$ for $D \in N_{def}^{\mathcal{T}_1}$, we can equivalently check wether $Z \sqsubseteq_{gfp,\mathcal{F},\mathcal{T}_4} X_D$, where $\mathcal{T}_4$ consists of a definition $X_D \equiv D$.

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

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

$$
\begin{aligned}
E &= \{\phi \mid \phi \in S \text{ and there is a } r \in N_{role} \text{ such that } \exists r.\phi \in S\}, \\
N_{pair} &= \{(X,Y) \mid X,Y \in (N_{def}^{\mathcal{T}} \cup E)\}, \text{ and} \\
G &= \{\phi \mid \phi \text{ is a subconcept over } N_{prim} \text{ and } N_{role}\}.
\end{aligned}
$$

Now we define the conservative extension of the TBox as follows.

Definition 4.1.4. 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

$$
(\phi,\psi) \equiv \theta_1 \sqcap \ldots \sqcap \theta_k \sqcap X_1 \sqcap \ldots \sqcap X_l \sqcap \exists r_1.(\phi_1,\psi_1) \sqcap \ldots \sqcap \exists r_m.(\phi_m,\psi_m)
$$

for each $(\phi,\psi) \in N_{pair}$, where:

1. $\theta \in \{\theta_1, ..., \theta_k\}$ iff $\phi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} \theta$, $\psi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} \theta$, and $\theta \in G$;

2. $X \in \{X_1, ..., X_l\}$ iff $\phi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} X$, $\psi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} X$, and $X \in N_{def}^{\mathcal{T}}$;

3. $\exists r.(\tau,\sigma) \in$ iff $\phi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} \exists r.\tau$, $\psi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} \exists r.\sigma$, $\{\exists r_1.(\phi_1,\psi_1), ..., \exists r_m.(\phi_m,\psi_m)\}$ and $(\tau,\sigma) \in N_{pair}$.

Least common subsumer of two defined concepts X and Y occurring in the TBox will be newly defined concept (X,Y). Notice that 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 $O((|\mathcal{F}|+|\mathcal{T}|)^2 * s)$ steps, where s is total number of pairs $(\phi,\psi) \in \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})}$. The algorithm iterates over all combinations for ϕ and ψ and checks $\sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})}$ relation for all θ such that $\phi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} \theta$ and $\psi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} \theta$, for all X such that $\phi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} X$ and $\psi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} X$, and for all $\exists r.\phi_1$ and $\exists r.\psi_1$ such that $\phi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} \exists r.\phi_1$ and $\psi \sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})} \exists r.\psi_1$. This time complexity is lower than the one of the algorithm for computing the $\sqsubseteq_{\infty}^{(\mathcal{F},\mathcal{T})}$ relation. Therefore, the overall

time complexity of computation of lcs is dominated by time complexity of the algorithm for deciding subsumption.

We will show that the definition above provides computation of lcs of two given concepts, i.e that the least common subsumer of two defined concepts X and Y occurring in the TBox is the newly defined concept (X, Y). What should be shown is that conditions 1. and 2. from the Definition 4.1.3 hold for all (X, Y). In order to do so, we will restrict our attention to those conservative extensions from the condition 2. for which newly added definitions are of a certain regular structure, i.e. we will require them to be *normalized modulo a TBox*. Notion of normalized terminologies is introduced in the next subsection.

4.1.1 Normalization of the terminologies

As announced, we will deal with the TBoxes fulfilling certain conditions, namely the condition of having *normalized* terminologies (or at least some parts of terminologies, and modulo a given TBox). Their fixed structure will provide easier way of showing properties of the TBoxes introduced in Definition 4.1.4.

As defined in [2], we say that the terminology part $\mathcal{T}$ of a hybrid TBox $(\mathcal{F}, \mathcal{T})$ is *normalized* if every definition in $\mathcal{T}$ is of the form:

$$X \equiv P_1 \sqcap P_2 \sqcap ... P_m \sqcap \exists r_1.X_1 \sqcap \exists r_2.X_2 \sqcap ... \sqcap \exists r_n.X_n.$$

Here, P_i is a primitive concept or the $\top$ symbol, for every i, while X and X_i are defined concepts for every i.

It has been shown in [2] that every terminology consisting of (cyclic) definitions can be transformed into an equivalent normalized one, which includes introducing some new concepts. Here we recall the normalization procedure given in [2] and modify it where necessary.

Assume now that $(\mathcal{F}, \mathcal{T})$ is a hybrid $\mathcal{EL}$ TBox, built over N_{prim}, N_{role} and N_{def}. Standard procedure for normalization of the terminology part of hybrid $\mathcal{EL}$ TBoxes consists of three steps. The procedure results in a TBox over N_{prim}, N_{role} and $N'_{def} \supseteq N_{def}$ which is normalized. Thus, when we say that a TBox is obtained by normalization of an another TBox, we mean it was obtained by applying the following:

- NT1: put $\mathcal{T}' = \mathcal{T}$ and $N'_{def} := N_{def}$. Repeat until no longer applicable: find a definition from $\mathcal{T}'$ of the form:

  $$X \equiv \phi_1 \sqcap ... \sqcap \phi_{k-1} \sqcap \exists r.\psi \sqcap \phi_{k+1} \sqcap ... \sqcap \phi_n,$$

  where ψ is not a defined concept from N'_{def}. If $\mathcal{T}'$ contains a definition of the form
  $$Z \equiv \psi,$$
  replace $X \equiv \phi_1 \sqcap ... \sqcap \phi_{k-1} \sqcap \exists r.\psi \sqcap \phi_{k+1} \sqcap ... \sqcap \phi_n$, with

  $$X \equiv \phi_1 \sqcap ... \sqcap \phi_{k-1} \sqcap \exists r.Z \sqcap \phi_{k+1} \sqcap ... \sqcap \phi_n.$$

Otherwise, add to $\mathcal{T}$ the definition

$$Z \equiv \psi,$$

and add Z to N'_{def}. Also, replace $X \equiv \phi_1 \sqcap ... \sqcap \phi_{k-1} \sqcap \exists r.\psi \sqcap \phi_{k+1} \sqcap ... \sqcap \phi_n$, with

$$X \equiv \phi_1 \sqcap ... \sqcap \phi_{k-1} \sqcap \exists r.Z \sqcap \phi_{k+1} \sqcap ... \sqcap \phi_n,$$

- NT2: identify disjoint classes C_X of mutually equivalent defined concepts in the following way: search the TBox for the sequences of the form $X = X_1, X_2, ..., X_{n-1}, X_n = X$, where X_{i+1} occurs as a conjunct on the top level in the definition of X_i. For such X_i and X_j, where $i \leq j$ we say that they are in $\sim$ relation. Relation $\sim$ is an equivalence relation on N_{def}, and C_X is the equivalence class containing X. Then, all X_i are mutually equivalent and belong to the same class.
  For each of the classes $C_Y = \{Y_1, ..., Y_m\}$, let D_Y be the set off all concept descriptions that occur as conjuncts at the top level in definitions of some of the elements of C_Y. Let also $\{\phi_1, ..., \phi_k\} = D_Y \setminus C_Y$ (therefore, ϕ_i is either in N'_{def} and not in C_Y, or in N_{prim}, or of the form $\exists r.Z$, with $Z \in N'_{def}$). Then, replace the definitions of all of the Y_i by

$$Y_i \equiv \phi_1 \sqcap ... \sqcap \phi_k.$$

- NT3: Proceed now as follows: repeat until there is no element from N'_{def} occurring as a conjunct at the top level of a definition form the TBox: find the defined concept V of N'_{def} for which no element of N'_{def} occurs in its definition on the top level (and such concept exists!) and replace all the occurrences of V in $\mathcal{T}$ by its definition.

Notice that NT1 serves to remove complex concept descriptions and primitive concepts from existential restrictions. It can be carried out in linear time in the size of the TBox, since the number of introductions of new definitions is bounded by the number of existential restrictions. Similarly, size of the new TBox is linear in the size of the original one, and NT1 yields a terminology with definitions that are conjunctions of primitive and defined concepts, and concept descriptions of the form $\exists r.W$, where W is defined concept.

Consider now so called *dependency graph of a TBox*, a directed graph with defined concepts as nodes such that there is an edge from X to Y iff Y occurs as a conjunct at top level conjunction of the definition of X. Finding classes of equivalence of the $\sim$ relation corresponds to finding strong components, i.e. subgraphs of the description graph such that there is a path between each two nodes of a subgraph. In addition, we require these subgraphs to be maximal, i.e., for every subgraph G it must hold that there is no other subgraph F of the description graph such that G is a subgraph of F and F is a connected

graph. It is known that finding strong components of a graph can be carried out in polynomial time in the size of the graph [28]. Computation of strong components allows polynomial runtime of the first part of NT2, computation of $\sim$ relation. As for the remaining part of NT2, computing the sets $D_Y \setminus C_Y$ requires polynomial time, since both, the size of D_Y and C_Y and the number of equivalence classes is bounded by the size of the TBox. Replacement of the definitions is, naturally of polynomial time complexity, since it is bounded by the number of defined concepts, and resulting TBox is of polynomial size in the size of the original TBox, since its growth is bounded by the number of defined concepts times the maximal size of $D_Y \setminus C_Y$, hence polynomial in the size of the original TBox.

After performing NT2, the dependency graph of the TBox is acyclic. NT3 is well defined and results in a normalized TBox since it corresponds to iterative removing of the node of the dependency graph with no edge starting in that node, until no nodes are left to be removed. A node with no edge starting in that node always have to exist in a cyclic graph. If we assume the opposite, starting from any node, we can always come to another node following an edge. At some point, we will arrive to a node that was already visited, since the number of nodes is finite. This means that the description graph has a cycle, which contradicts the assumption. After removing the node with no exit edges, remaining dependency graph is still acyclic, and a new node can be found with no exit edge. NT3 can be carried out in polynomial time, since the runtime is bounded by the number of definitions. NT3 can result in quadratic blowup of the TBox because every definition can be replaced with something that can be linear in the size of the TBox.

We saw that normalization requires polynomial time and the result is polynomial (quadratic) in the size of the original TBox. Soundness of the normalization procedure is shown in the following lemma.

Lemma 4.1.2. *Let $(\mathcal{F}, \mathcal{T})$ be a hybrid $\mathcal{EL}$ TBox, and $(\mathcal{F}, \mathcal{T}')$ the one obtained after normalization. Then every gfp model of $(\mathcal{F}, \mathcal{T})$ can be extended to a gfp model of $(\mathcal{F}, \mathcal{T}')$ and vice versa, every gfp model of $(\mathcal{F}, \mathcal{T}')$ can be restricted to a model of gfp model of $(\mathcal{F}, \mathcal{T})$.*

Proof. As usual, N_{prim}, N_{def} and N_{role} denote sets of primitive-, defined concepts and role names occurring in $(\mathcal{F}, \mathcal{T})$, respectively, while primitive-, defined concepts and role names occurring in $(\mathcal{F}, \mathcal{T}')$ will be denoted by N_{prim}, N'_{def} and N_{role}. Clearly, $N_{def} \subseteq N'_{def}$.

In order to prove the claim, we may choose a model of $(\mathcal{F}, \mathcal{T})$ and then try to find its extension to a model of $(\mathcal{F}, \mathcal{T}')$. Also, in that case, we should choose a model of $(\mathcal{F}, \mathcal{T}')$ and try to show that it can be restricted to a model of $(\mathcal{F}, \mathcal{T})$.

Notice that a gfp model of a hybrid TBox is determined by its restriction to primitive concepts and role names. Therefore, alternatively to the previous

idea, we may try to show that once we fix a primitive interpretation that is a model of $\mathcal{F}$, corresponding gfp models of $(\mathcal{F}, \mathcal{T})$ and $(\mathcal{F}, \mathcal{T}')$ will be the required models, i.e. the model of $(\mathcal{F}, \mathcal{T}')$ will be an extension of the model of $(\mathcal{F}, \mathcal{T})$ and the later will be a restriction of the former one.

Let $\mathcal{I}$ be the gfp model of $(\mathcal{F}, \mathcal{T})$ and $\mathcal{I}'$ be the gfp model of $(\mathcal{F}, \mathcal{T}')$, such that $\mathcal{I}$ and $\mathcal{I}'$ coincide on elements of N_{prim} and N_{role}. Thus, it suffice to show that $\mathcal{I}$ and $\mathcal{I}'$ coincide on elements of N_{def}.

To that purpose, we extend the interpretation $\mathcal{I}$ to an interpretation $\mathcal{I}_1$ of $(\mathcal{F}, \mathcal{T}')$ as follows: let $\{Z_1, ..., Z_n\}$ be all the elements of N'_{def} introduced in NT1, such that Z_i does not occur as a subconcept in the definition of Z_j, for $j > i$. Notice that this is always possible by the way NT1 is carried out. Then, we extend $\mathcal{I}_1$ by $\mathcal{I}_1(Z_i) = \mathcal{I}_1(\phi_{Z_i})$, for $i = 1, ..., n$, where ϕ_{Z_i} is definition of Z_i after NT1. Clearly, $\mathcal{I}_1$ is a descriptive model of $\mathcal{T}$ after NT1. We show now that $\mathcal{I}_1(X) \subseteq \mathcal{I}_1(\phi_X)$ for every definition $X \equiv \phi_X$ after the steps NT2 and NT3. As a descriptive model of the terminology prior to NT2, $\mathcal{I}_1$ clearly interprets the elements of a cycle in the dependency graph as equal, and so does the elements of the same class C_X. But then, $\mathcal{I}_1$ interprets X as a subset of $\mathcal{I}_1(Z)$, where Z is an arbitrary subconcept occurring as a conjunct at the top level of definitions of some concept form C_X prior to NT2. Hence, $\mathcal{I}_1(X) \subseteq \mathcal{I}_1(\phi_X)$ for every definition $X \equiv \phi_X$ after the step NT2. NT3 does not affect this property, since we replace right-hand side of the definitions by the concept names that are interpreted by $\mathcal{I}_1$ as larger sets.

To conclude, $\mathcal{I}_1$ is an extension of $\mathcal{I}$ that fulfills conditions (1)-(5) of Definition 2.2.6 applied to $\mathcal{T}'$ and its gfp model $\mathcal{I}'$. As a consequence, $\mathcal{I}(X) = \mathcal{I}_1(X) \subseteq \mathcal{I}'(X)$ for all $X \in N_{def}$.

On the other hand, $\mathcal{I}'$ is also a (descriptive) model of the original TBox. Indeed, notice that if $\mathcal{I}'$ is a model of the TBox, then it was also a (descriptive) model of the TBox prior to NT3 step of normalization, since NT3 replaces definitions by other ones that are interpreted by $\mathcal{I}'$ as the same sets. As for the NT2 step, $\mathcal{I}'$ is clearly a model of all of the definitions of the TBox prior to NT2, since those definitions are conjunctions of either: a subconcept that is a conjunct at the top level from the definition after NT2, or a defined concept that $\mathcal{I}'$ interprets as equal. Finally, NT1, as NT3, replaces equals for equals from the point of $\mathcal{I}'$, therefore $\mathcal{I}'$ is a model of $(\mathcal{F}, \mathcal{T})$. In particular, $\mathcal{I}'$ fulfills conditions (1)-(5) of Definition 2.2.6 applied to $\mathcal{T}$ and its gfp model $\mathcal{I}$. This implies $\mathcal{I}_1(X) \subseteq \mathcal{I}(X)$ for all $X \in N_{def}$. $\qquad\square$

The following is immediate.

Corollary 4.1.1. *For all subconcepts ϕ and ψ occurring in the $(\mathcal{F}, \mathcal{T})$, it holds $(\mathcal{F}, \mathcal{T}) \models \phi \sqsubseteq \psi$ iff $(\mathcal{F}', \mathcal{T}') \models \phi \sqsubseteq \psi$.*

After showing some of the properties of the normalization, namely that it preserves the subsumption relation and that, to a certain extent, it preserves

the models, we proceed by considering conservative extensions of the TBoxes, augmented by a variant of normalized definitions.

4.1.2 Conservative extensions with normalized definitions

In order to simplify future discussion on the correctness of the definition of the lcs given in Definition 4.1.4, we will restrict our attention to a subset of all possible conservative extensions from Definition 4.1.3.

Definition 4.1.5. 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 from $\mathcal{T}' \setminus \mathcal{T}$ is of the form:

$$Z \equiv P_1 \sqcap ... \sqcap P_m \sqcap X_1 \sqcap ... \sqcap X_k \sqcap \exists r_1.Z_1 \sqcap ... \sqcap \exists r_n.Z_n$$

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

Notice that $(\mathcal{F}, \mathcal{T}')$ from the definition above is not normalized as described in the previous section. However, if we treat concepts defined in $\mathcal{T}$ as primitive concepts in definitions from $\mathcal{T}' \setminus \mathcal{T}$, those definitions are indeed normalized as described in previous section. This is also the reason for using the term *normalized definitions modulo* $(\mathcal{F}, \mathcal{T})$.

The following proposition and the lemmas that precede it show that the property 2. of the Definition 4.1.3 can be, without loss of generality, restricted to only those conservative extensions of $(\mathcal{F}, \mathcal{T}_{lcs})$ that are obtained by adding the normalized definitions modulo the TBox.

Lemma 4.1.3. *For any conservative extension $(\mathcal{F}, \mathcal{T} \cup \mathcal{B})$ of a hybrid $\mathcal{EL}$ TBox $(\mathcal{F}, \mathcal{T})$, every gfp model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{B})$ is also a gfp model of $\mathcal{B}$, where we consider concepts defined in $\mathcal{T}$ as primitive in $\mathcal{B}$ when considering gfp models of $\mathcal{B}$.*

Proof. Let N_{prim}, N_{role} and N_{def} denote, the set of primitive concepts, role names and the set of defined concepts in $(\mathcal{F}, \mathcal{T})$, respectively. We denote the set of concepts defined in $\mathcal{B}$ by $N_{def}^{\mathcal{B}}$.

A gfp model $\mathcal{I}$ of $(\mathcal{F}, \mathcal{T} \cup \mathcal{B})$ clearly has to be a descriptive model of $\mathcal{B}$. Assume now that there is a function $\mathcal{J}$ satisfying (1)-(5) form Definition 2.2.6 applied to the TBox $\mathcal{B}$ and to the model $\mathcal{I}$. We show $\mathcal{J}(\phi) \subseteq \mathcal{I}(\phi)$ for all ϕ over N_{prim}, N_{role}, N_{def} and $N_{def}^{\mathcal{B}}$. Indeed, if $\mathcal{J}$ satisfies (1)-(5) form Definition 2.2.6 applied to the TBox $\mathcal{B}$ and $\mathcal{I}$, then $\mathcal{J}$ also satisfies (1)-(5) form Definition 2.2.6 applied to the TBox $\mathcal{T} \cup \mathcal{B}$ and $\mathcal{I}$. Conditions (1), (4) and (5) are immediately satisfied. For (2), it suffice to notice that since $\mathcal{J}(W) = \mathcal{I}(W)$, for $W \in N_{prim} \cup N_{def}$, then $\mathcal{J}(W) = \mathcal{I}(W)$, for $W \in N_{prim}$. For (3), it suffice to notice that $\mathcal{J}(X) \subseteq \mathcal{J}(\phi_X)$ for

$X \in N^{\mathcal{B}}_{def}$, while for $X \in N_{def}$ it can be noticed that $\mathcal{J}(X) = \mathcal{I}(X)$ and $\mathcal{J}(\phi_X) = \mathcal{I}(\phi_X)$, since $\mathcal{J}$ coincides with $\mathcal{I}$ on N_{prim}, N_{role} and N_{def}. This further implies $\mathcal{J}(X) = \mathcal{I}(X) = \mathcal{I}(\phi_X) = J(\phi_X)$, for $X \in N_{def}$. To conclude, $\mathcal{J}$ satisfies (1)-(5) form Definition 2.2.6 applied to the TBox $\mathcal{T} \cup \mathcal{B}$ and $\mathcal{I}$, and since $\mathcal{I}$ is a gfp model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{B})$, $\mathcal{J}(\phi) \subseteq \mathcal{I}(\phi)$ holds for all ϕ over N_{prim}, N_{role}, N_{def} and $N^{\mathcal{B}}_{def}$, i.e. $\mathcal{I}$ is a gfp model of $\mathcal{B}$.

$\square$

Lemma 4.1.4. *Every gfp model $\mathcal{I}$ of a conservative extension $(\mathcal{F}, \mathcal{T} \cup \mathcal{B})$ of the hybrid $\mathcal{EL}$ TBox $(\mathcal{F}, \mathcal{T})$ is an extension of a gfp model of $(\mathcal{F}, \mathcal{T})$.*

Proof. Again, let N_{prim}, N_{role} and N_{def} denote, the set of primitive concepts, role names and the set of defined concepts in $(\mathcal{F}, \mathcal{T})$, respectively. We denote the set of concepts defined in $\mathcal{B}$ by $N^{\mathcal{B}}_{def}$.

A gfp model $\mathcal{I}$ of $(\mathcal{F}, \mathcal{T} \cup \mathcal{B})$ clearly has to be a descriptive model of $(\mathcal{F}, \mathcal{T})$. Assume now that there is a function $\mathcal{J}$ satisfying (1)-(5) form Definition 2.2.6 applied to the TBox $\mathcal{T}$ and to the model $\mathcal{I}$. We show $\mathcal{J}(\phi) \subseteq \mathcal{I}(\phi)$ for all ϕ over N_{prim}, N_{role}, and N_{def}. Indeed, if $\mathcal{J}$ satisfies (1)-(5) form Definition 2.2.6 applied to the TBox $\mathcal{T}$ and $\mathcal{I}$, then $\mathcal{J}$ also satisfies (1)-(5) form Definition 2.2.6 applied to the TBox $\mathcal{T} \cup \mathcal{B}$ and $\mathcal{I}$. Conditions (1), (2), (4) and (5) are immediately satisfied, while for (3) follows from the $\mathcal{T} \subseteq \mathcal{T} \cup \mathcal{B}$. $\square$

Proposition 4.1.1. *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 $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1) \models \phi \sqsubseteq \psi$ iff $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2) \models \phi \sqsubseteq \psi$ for every two concepts ϕ and ψ defined in $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$.*

Proof. Let N_{prim}, N_{role} and N_{def} denote, as before, the sets of primitive concepts, role names and the set of defined concepts in $(\mathcal{F}, \mathcal{T})$. We denote the set of concepts defined in $\mathcal{A}_1$ by $N^{\mathcal{A}_1}_{def}$. Let $\mathcal{A}_2$ be the TBox obtained by applying the normalization algorithm to $\mathcal{A}_1$, where we treat elements of N_{prim} and N_{def} as primitive concepts in $\mathcal{A}_1$, i.e. we consider only elements of $N^{\mathcal{A}_1}_{def}$ to be defined concepts while performing the normalization steps NT1, NT2 and NT3. Notice that definitions in $\mathcal{A}_2$ are normalized modulo $(\mathcal{F}, \mathcal{T})$.

Every gfp model $\mathcal{I}$ of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$ is a gfp model of $\mathcal{A}_1$ by Lemma 4.1.3. By Lemma 4.1.2 (with empty foundation part of the TBox, treating the elements of N_{def} as primitive), $\mathcal{I}$ can be extended to $\mathcal{J}$, a gfp model of $\mathcal{A}_2$. Now, $\mathcal{J}$ is a model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$. Indeed, a gfp model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$ has

to be a gfp model of $(\mathcal{F}, \mathcal{T})$, by Lemma 4.1.4, therefore, it has to coincide with $\mathcal{I}$ (also $\mathcal{J}$) on N_{prim}, N_{role} and N_{def}. A gfp model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$, once fixed on N_{prim}, N_{role} and N_{def}, has to be a gfp model of $\mathcal{A}_2$ where concepts of N_{def} are treated as primitive, according to Lemma 4.1.3. Such model is unique once the interpretation of elements of N_{prim}, N_{role} and N_{def} is fixed, so it has to coincide with $\mathcal{J}$ on elements of $N_{def}^{\mathcal{A}_2}$, since $\mathcal{J}$ is also a gfp model of $\mathcal{A}_2$ that coincides with the gfp model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$ on N_{prim}, N_{role} and N_{def}.

Vice versa, every model of $\mathcal{J}$ of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$ is gfp model of $\mathcal{A}_2$ by Lemma 4.1.3. By Lemma 4.1.2 (with empty foundation part of the TBox), $\mathcal{J}$ can be restricted to $\mathcal{I}$, a gfp model of $\mathcal{A}_1$. Again, $\mathcal{I}$ is a model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$. Indeed, a gfp model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$ has to be a gfp model of $(\mathcal{F}, \mathcal{T})$, by Lemma 4.1.4, therefore, it has to coincide with $\mathcal{J}$ (also $\mathcal{I}$) on N_{prim}, N_{role} and N_{def}. A gfp model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$, once fixed on N_{prim}, N_{role} and N_{def}, has to be a gfp model of $\mathcal{A}_1$ where concepts of N_{def} are treated as primitive, according to Lemma 4.1.3. Such model is unique once the interpretation of elements of N_{prim}, N_{role} and N_{def} is fixed, so it has to coincide with $\mathcal{I}$ on elements of $N_{def}^{\mathcal{A}_1}$, since $\mathcal{I}$ is also a gfp model of $\mathcal{A}_1$ that coincides with the gfp model of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$ on N_{prim}, N_{role} and N_{def}.

In particular, the previous means that $\phi \sqsubseteq \psi$ holds under all gfp models of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$, if an only iff it holds under all gfp models of $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$.

$\square$

Previous proposition, as announced before, shows that one can restrict the attention to the conservative extensions obtained by adding the normalized definitions modulo a TBox when checking for the 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 X and Y. By 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 (X, Y) will be subsumed by Φ w.r.t. $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_1)$ iff (X, Y) is subsumed by Φ w.r.t. $(\mathcal{F}, \mathcal{T} \cup \mathcal{A}_2)$. In particular, (X, Y) will be subsumed by all of the Φs that subsume both X and Y w.r.t. an arbitrary conservative extension of the TBox iff it is subsumed by all of the Φs that subsume both X and Y w.r.t. conservative extensions of the TBox obtained by adding normalized definitions modulo $(\mathcal{F}, \mathcal{T})$. This exactly means that it suffice to consider only the conservative extensions obtained by adding the normalized definitions modulo a TBox when checking for the property 2.

4.2   Correctness of the procedure

This section is dedicated to proving that (X, Y) form Definition 4.1.4 is indeed the least-common subsumer of defined concepts X and Y occurring in the $\mathcal{EL}$ TBox $(\mathcal{F}, \mathcal{T})$.

In order to prove that (X, Y) is the lcs of X and Y, one has to show conditions 1. and 2. from the definition of hybrid lcs. This will be done in two steps. In both cases, soundness and completeness of HC will play crucial role.

We begin by showing one property of the $\sqsubseteq_n$ and $\sqsubseteq_\infty$ relations.

Lemma 4.2.1. *Let ϕ, ψ and θ be arbitrary subconcepts occurring in a TBox $(\mathcal{F}, \mathcal{T})$. If $\phi \sqsubseteq_\infty \theta$ and $\theta \sqsubseteq_n \psi$, then $\phi \sqsubseteq_n \psi$.*

Proof. Proof is done by induction on derivation of $\theta \sqsubseteq_n \psi$. We distinguish cases depending on the last rule applied. Base of the induction are the cases where $\theta \sqsubseteq_n \psi$ is obtained by applying (Ax), (Top) or (Start).

- (Ax): In this case $\theta = \psi$, thus $\phi \sqsubseteq_n \psi$ follows from $\phi \sqsubseteq_\infty \psi$;

- (Top): In this case $\psi = \top$, thus $\phi \sqsubseteq_n \psi$ follows by applying (Top);

- (Start): In this case $n = 0$, thus $\phi \sqsubseteq_n \psi$ follows by applying (Start);

- (AndLi), for $i = 1, 2$: In this case $\theta = \theta_1 \sqcap \theta_2$, and $\theta_1 \sqcap \theta_2 \sqsubseteq_n \psi$ follows from $\theta_i \sqsubseteq_n \psi$. By completeness of HC, $\phi \sqsubseteq_\infty \theta_1 \sqcap \theta_2$ implies $\phi \sqsubseteq_\infty \theta_i$, thus induction hypothesis can be applied and it yields $\phi \sqsubseteq_n \psi$;

- (AndR): In this case $\psi = \psi_1 \sqcap \psi_2$, and $\theta \sqsubseteq_n \psi_1 \sqcap \psi_2$ follows from $\theta \sqsubseteq_n \psi_1$ and $\theta \sqsubseteq_n \psi_2$. Induction hypothesis can be applied to the premises and it yields $\phi \sqsubseteq_n \psi_1$ and $\phi \sqsubseteq_n \psi_2$. Now, $\phi \sqsubseteq_n \psi_1 \sqcap \psi_2$ follows by applying (AndR);

- (DefR): In this case $\theta \sqsubseteq_n \psi$ follows from $\theta \sqsubseteq_{n-1} \phi_\psi$, where $\psi \equiv \phi_\psi$ is a definition from $\mathcal{T}$. Induction hypothesis can be applied and it yields $\phi \sqsubseteq_{n-1} \phi_\psi$, thus $\phi \sqsubseteq_n \psi$ follows by applying (DefR);

- (DefL): In this case $\theta \sqsubseteq_n \psi$ follows from $\phi_\theta \sqsubseteq_n \psi$, where $\theta \equiv \phi_\theta$ is a definition from $\mathcal{T}$. By completeness of HC, $\phi \sqsubseteq_\infty \phi_\theta$. Induction hypothesis can be applied and it yields $\phi \sqsubseteq_n \psi$;

- (Concept): In this case $\theta \sqsubseteq_n \psi$ follows from $\theta \sqsubseteq_n \alpha$ and $\beta \sqsubseteq_n \psi$, where $\alpha \sqsubseteq \beta$ is a GCI from $\mathcal{F}$. Induction hypothesis can be applied and it yields $\phi \sqsubseteq_n \alpha$, thus $\phi \sqsubseteq_n \psi$ follows by applying (Concept);

- (Ex): In this case $\theta = \exists r.\theta_1$ and $\psi = \exists r.\psi_1$, and $\exists r.\theta_1 \sqsubseteq_n \exists r.\psi_1$ follows from $\theta_1 \sqsubseteq_n \psi_1$. By Lemma 3.1.4 there exist α, β and ρ such that $\phi \sqsubseteq_\infty \alpha$, $\beta \Vdash \exists r.\rho$ and $\rho \sqsubseteq_\infty \theta_1$, where $\alpha \sqsubseteq \beta$ is a GCI in $\mathcal{F}$, or $\phi = \alpha = \beta$. If $\alpha \sqsubseteq \beta$ is a GCI in $\mathcal{F}$, induction hypothesis can be applied to $\rho \sqsubseteq_\infty \theta_1$ and $\theta_1 \sqsubseteq_n \psi_1$, and it yields $\rho \sqsubseteq_n \psi_1$. Applying (Ex) to $\rho \sqsubseteq_n \psi_1$ yields $\exists r.\rho \sqsubseteq_n \exists r.\psi_1$. By Lemma 3.1.1, applied to $\beta \Vdash \exists r.\rho$ and $\exists r.\rho \sqsubseteq_n \exists r.\psi_1$ yields $\beta \sqsubseteq_n \exists r.\psi_1$, thus $\phi \sqsubseteq_n \exists r.\psi_1$ follows by applying (Concept).
  If $\phi = \alpha = \beta$, Lemma 3.1.1 can be applied to $\phi \Vdash \exists r.\rho$ and $\exists r.\rho \sqsubseteq_n \exists r.\psi_1$ to obtain $\phi \sqsubseteq_n \exists r.\psi_1$.

$\square$

Lemma 4.2.2. *Let ϕ and ψ be arbitrary concepts from $N_{def}^{\mathcal{T}} \cup E$. Then, $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi, \psi)$ and $\psi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi, \psi)$ for every n.*

Proof. We give a proof of $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi, \psi)$, proof of $\psi \sqsubseteq_n (\phi, \psi)$ is analogous.

Proof is carried out by induction on n.

For $n = 0$, $\phi \sqsubseteq_0^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi, \psi)$ follows from the rule (Start).

Assume now that $\phi \sqsubseteq_l^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi, \psi)$ holds for all $l \leq n$. We prove that $\phi \sqsubseteq_{n+1}^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi, \psi)$. Let

$$(\phi, \psi) \quad \equiv \quad \theta_1 \sqcap \ldots \sqcap \theta_k \sqcap X_1 \sqcap \ldots \sqcap X_u \sqcap \exists s_1.(\phi_{l_1}, \psi_{m_1}) \sqcap \ldots \sqcap \exists s_t.(\phi_{l_t}, \psi_{m_t})$$

be the definition of (ϕ, ψ) in the extended hybrid TBox $(\mathcal{F}, \mathcal{T}_{lcs})$. One of the properties of the $\sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})}$ relation, shown in Lemma 3.1.2, is that $\xi \sqsubseteq_{n+1}^{(\mathcal{F}, \mathcal{T}_{lcs})} Z$ iff $\xi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \phi_Z$, where $Z \equiv \phi_Z$ is a definition from $\mathcal{T}_{lcs}$. In this case, $\phi \sqsubseteq_{n+1}^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi, \psi)$ iff $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \theta_1 \sqcap \ldots \sqcap \theta_k \sqcap X_1 \sqcap \ldots \sqcap X_u \sqcap \exists s_1.(\phi_{l_1}, \psi_{m_1}) \sqcap \ldots \sqcap \exists s_t.(\phi_{l_t}, \psi_{m_t})$. Therefore, it suffice to show $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \theta_1 \sqcap \ldots \sqcap \theta_k \sqcap X_1 \sqcap \ldots \sqcap X_u \sqcap \exists s_1.(\phi_{l_1}, \psi_{m_1}) \sqcap \ldots \sqcap \exists s_t.(\phi_{l_t}, \psi_{m_t})$. One way to show this is to give a prove of $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \theta_i$ for $i = 1, \ldots, k$, $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} X_i$ for $i = 1, \ldots, u$, and $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \exists s_j.(\phi_{l_j}, \psi_{m_j})$ for $j = 1, \ldots, t$. Then, $\phi \sqsubseteq_{n+1}^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi, \psi)$ will follow by applying (AndR) several times, and (DefR) in the end.

- $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \theta_i$: by Definition 4.1.4, $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_{lcs})} \theta_i$. Therefore, by definition of $\sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_{lcs})}$, $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \theta_i$, for $i = 1, \ldots, k$.

- $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} X_i$: by Definition 4.1.4, $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_{lcs})} X_i$. Therefore, by definition of $\sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_{lcs})}$, $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} X_i$, for $i = 1, \ldots, k$.

- $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \exists s_j.(\phi_{l_j}, \psi_{m_j})$: by Definition 4.1.4, $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_{lcs})} \exists s_j.\phi_{l_j}$, therefore $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \exists s_j.\phi_{l_j}$. Since both ϕ_{l_j} and ψ_{m_j} belong to the $N_{def} \cup E$, induction hypothesis can be applied and it yields $\phi_{l_j} \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} (\phi_{l_j}, \psi_{m_j})$. Then, $\exists s_j.\phi_{l_j} \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \exists s_j.(\phi_{l_j}, \psi_{m_j})$ follows by applying the rule (Ex). Since $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_{lcs})} \exists s_j.\phi_{l_j}$, Lemma 4.2.1 yields $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_{lcs})} \exists s_j.(\phi_{l_j}, \psi_{m_j})$.

$\square$

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

We show now another, rather technical property of conservative extensions and $\Vdash^{(\mathcal{F}, \mathcal{T})}$ relation that will be used in later proofs. Here $\sigma \Vdash^{(\mathcal{F}, \mathcal{T})} \tau$ is, as

before, used to denote that judgement $\sigma \sqsubseteq_n^{(\mathcal{F},\mathcal{T})} \tau$ can be derived using rules (Ax), (AndL1), (AndL2), (DefL) and (Concept), for some n.

Lemma 4.2.3. *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 $(\mathcal{F},\mathcal{T})$.*

Proof. Proof is done by induction on derivation of $\sigma \Vdash^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\tau$. If $\sigma \Vdash^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\tau$ is derived using the rule (Ax), $\exists r.\tau$ is a subconcept occurring in $(\mathcal{F},\mathcal{T})$ by assumption. Suppose now the last rule applied in the derivation of $\sigma \Vdash^{(\mathcal{F},\mathcal{T}_2)} \exists r.\tau$ is one of the (AndL1), (AndL2), (DefL) or (Concept). In all four cases induction hypothesis can be applied to the premisses of the rules and it will yields the claim of the lemma. $\qquad\square$

Lemma 4.2.4. *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 ϕ and ψ be two concepts from $N_{def}^{\mathcal{T}} \cup E$, and let Φ be a concept defined in $\mathcal{T}_2 \setminus \mathcal{T}$. If $\phi \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$ and $\psi \sqsubseteq_\infty^{(\mathcal{F},\mathcal{T}_2)} \Phi$, then $(\phi,\psi) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \Phi$, 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 G (form 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 $\oplus$ 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 $\oplus$ is defined in $\mathcal{T}_{lcs}$).

Proof is carried out by induction on n.

For $n = 0$, $(\phi,\psi) \sqsubseteq_0^{(\mathcal{F},\mathcal{T}_2)} \Phi$ follows from the rule (Start).

Assume now that $(\phi,\psi) \sqsubseteq_k^{(\mathcal{F},\mathcal{T}_2)} \Phi$ for all $k \leq n$. We prove that $(\phi,\psi) \sqsubseteq_{n+1}^{(\mathcal{F},\mathcal{T}_2)} \Phi$. One of the properties of the $\sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})}$ relation, shown in Lemma 3.1.2, is that $\xi \sqsubseteq_{n+1}^{(\mathcal{F},\mathcal{T}_{lcs})} Z$ iff $\xi \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_{lcs})} \Phi_Z$, where $Z \equiv \Phi_Z$ is a definition from $\mathcal{T}_{lcs}$. In our case, $(\phi,\psi) \sqsubseteq_{n+1}^{(\mathcal{F},\mathcal{T}_2)} \Phi$ iff $(\phi,\psi) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_u \sqcap \exists r_1.\Phi_1 \sqcap ... \sqcap \exists r_l.\Phi_l$. Therefore, it suffice to prove $(\phi,\psi) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \theta_1 \sqcap ... \sqcap \theta_k \sqcap X_1 \sqcap ... \sqcap X_u \sqcap \exists r_1.\Phi_1 \sqcap ... \sqcap \exists r_l.\Phi_l$. Again, one way to show this is to give a proof of $(\phi,\psi) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \theta_i$ for $i = 1,...,k$, $(\phi,\psi) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} X_i$ for $i = 1,...,u$, and $(\phi,\psi) \sqsubseteq_n^{(\mathcal{F},\mathcal{T}_2)} \exists r_j.\Phi_j$ for $j = 1,...,l$. Then, $(\phi,\psi) \sqsubseteq_{n+1} \Phi$ will follow by applying (AndR) several several times, and (DefR) in the end. Notice that the assumption on normalized definition modulo $(\mathcal{F},\mathcal{T}_{lcs})$ of Φ simplified the discussion to the point that it suffice to show the following:

48

- $(\phi, \psi) \sqsubseteq_n \theta_i$: by soundness and completeness of HC, $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \Phi$ implies $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \theta_i$, similarly, $\psi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \theta_i$, and therefore θ_i occurs on the right-hand side of the definition of (ϕ, ψ) by Definition 4.1.4, since θ_i belongs to G. Therefore, $(\phi, \psi) \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_2)} \theta_i$ follows from completeness of the HC calculus and the fact that $(\phi, \psi) \sqsubseteq \theta_i$ holds in all models of $(\mathcal{F}, \mathcal{T})$.

- $(\phi, \psi) \sqsubseteq_n X_i$: we distinguish two cases

  1. X_i is defined in $\mathcal{T}$: by soundness and completeness of HC, $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)}$ Φ implies $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} X_i$, similarly, $\psi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} X_i$, and therefore X_i occurs on the right-hand side of the definition of (ϕ, ψ) by Definition 4.1.4. Thus, $(\phi, \psi) \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_2)} X_i$ follows from completeness of the HC calculus and the fact that $(\phi, \psi) \sqsubseteq X_i$ 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, $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \Phi$ implies $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)}$ (γ, δ), similarly, $\psi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} (\gamma, \delta)$. Now, 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 $(\phi, \psi) \sqsubseteq_n (\gamma, \delta)$.

- $(\phi, \psi) \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_2)} \exists r_j.\Phi_j$: again, by soundness and completeness of HC, $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \Phi$ implies $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \exists r_j.\Phi_j$ and $\psi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \Phi$ implies $\psi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \exists r_j.\Phi_j$. By Lemma 3.1.4, this means that there exist concepts α, β and ρ and such that

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

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

This further implies $\phi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \exists r_j.\rho$ by applying rule (Concept) to $\phi \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_2)} \alpha$ and $\beta \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_2)} \exists r_j.\rho$ for every n. In the same way we conclude $\psi \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \exists r_j.\rho_1$.

Lemma 4.2.3 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 E.

Now, induction hypothesis can be applied to $\rho \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)} \Phi_j$ and $\rho_1 \sqsubseteq_\infty^{(\mathcal{F}, \mathcal{T}_2)}$ Φ_j to obtain $(\rho, \rho_1) \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_2)} \Phi_j$. On the other hand, by Definition 4.1.4, $\exists r_j.(\rho, \rho_1)$ is one of the conjuncts in definition of (ϕ, ψ). Now, $(\phi, \psi) \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_2)} \exists r_j.\Phi_j$ can be derived form $(\rho, \rho_1) \sqsubseteq_n^{(\mathcal{F}, \mathcal{T}_2)} \Phi_j$, by applying (Ex) rule, (AndL1) or (AndL2) rules several times and (DefL) in the end.

$\square$

Again, due to the soundness of derivations in HC, considering defined concepts X, Y and the corresponding (X, Y), we have that (X, Y) is subsumed by every concept defined in $\mathcal{T}_2 \setminus \mathcal{T}$ that subsumes both X and Y. (We use notation from previous lemma.) By the comment after Definition 4.1.3, this conclusion is sufficient to show property 2. form 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. form the definition of hybrid lcs, this proves the following theorem.

Theorem 4.2.1. *The concept description (X, Y) form the extended hybrid TBox $(\mathcal{F}, \mathcal{T}_{lcs})$ is least common subsumer of X and Y w.r.t. the hybrid TBox $(\mathcal{F}, \mathcal{T})$.*

Chapter 5

Conclusions, related and future work

In this thesis, we considered the subsumption problem in description logic $\mathcal{EL}$ w.r.t. hybrid TBoxes. The main task was to check whether a proof-theoretic decision procedure for the subsumption problem for the case of hybrid TBoxes can be obtained by combining a proof-theoretic decision procedure for the case of GCIs interpreted by descriptive semantics with a procedure for the TBoxes containing cyclic definitions interpreted by greatest fixpoint semantics, given in [20].

We gave a positive answer to this question by devising a sound and complete calculus for deciding subsumption in $\mathcal{EL}$ w.r.t. hybrid TBoxes. Proofs of the soundness and completeness required detailed analysis of the calculus. We identified a provability relation in the calculus with the subsumption relation. A polynomial runtime decision procedure based on the proof search in the calculus was obtained. Two important points that enabled the polynomial runtime of the procedure were the fact that it suffices to consider only the subconcepts occurring in the TBox in order to prove a subsumption, and the fact that the rules of the calculus are such that they facilitate a polynomial bottom-up proof search.

An implemented subsumption reasoner Hyb for $\mathcal{EL}$ w.r.t. hybrid TBoxes, based on the mentioned decision procedure, is described in [23]. The implemented reasoner can also be applied for classifying ontologies consisting of GCIs interpreted by descriptive semantics, or for deciding subsumption for TBoxes consisting of cyclic definitions interpreted by greatest fixpoint semantics, as special cases of hybrid TBoxes. Also, it can be used for ontologies consisting of definitions where the descriptive and greatest fixpoint semantics coincide. For that reason, Hyb can be compared with the existing reasoners such as CEL (http://lat.inf.tu-dresden.de/systems/cel/, [6], [7], [5], [31], [13], [12]) for descriptive semantics, and a subsumption reasoner for terminologies interpreted

by greatest fixpoint semantics, given in [30]¹ on large fragments of life science ontologies such as GALEN. Another existing reasoner is a subsumption reasoner $\mathcal{EL}$ w.r.t. hybrid TBoxes given in [21], which is based on the gfp reasoner from [30].

Performance of the reasoner based on our decision procedure is, as expected, considerably worse on the fragments of GALEN than the performance of the CEL reasoner, since the CEL reasoner is well suited for the descriptive semantics induced by these life science test ontologies. When compared to a gfp reasoner from [30] or an existing reasoner described in [21] for deciding subsumption in $\mathcal{EL}$ w.r.t. hybrid semantics, the efficiency of the Hyb reasoner is, somewhat surprisingly, considerably better than the efficiency of the two existing reasoners. In the case of the reasoner w.r.t. the hybrid TBoxes, this can perhaps be explained by the better worst-case complexity of the decision procedure. (See [23] for details).

Another issue that was addressed in this thesis is the problem of computing the least common subsumers. The existing algorithms from [8] and [15] require normalized TBoxes. One shortcoming of normalization (besides the quadratic blowup of the size of the TBox) is the fact that if normalization is performed on a TBox prior to computation of the lcs, computation of the lcs will yield a TBox that is not a conservative extension of the original TBox. In this thesis we gave a novel algorithm for computation of the lcs in the case of arbitrary $\mathcal{EL}$ hybrid TBoxes. While the computation itself does not have to be carried out using a devised proof-theoretic technique, (it can be carried out by any tool capable of deciding subsumption w.r.t. hybrid TBoxes), the proof of correctness of the algorithm is essentially proof-theoretic.

There are several natural questions that arise at this point, that could result in future work extending the results given here. One could ask, for instance, if it is possible to have a similar, hopefully efficient, proof-theoretic treatment of extensions of $\mathcal{EL}$ where subsumption is known to be tractable, and if it is possible to apply a similar techniques outside of the $\mathcal{EL}$ family. Another question is whether a meaningful and efficient treatment of the knowledge bases including ABoxes could be given. In particular, one might investigate the possibility of having a proof-theoretic computation of the most specific concept as an inference service that, combined with the least common subsumer, facilitates the bottom-up building of the ontologies [16]. Some of those questions were addressed in [20], and the others are subject to the ongoing investigations.

---

¹Both reasoners were developed at Technische Universität Dresden, the first one by Baader, Lutz, Suntisrivaraporn, et.al., and the later one was a part of the Master's thesis of B. Suntisrivaraporn

Contents

Bibliography

[1] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat Genet*, 25(1):25–29, May 2000.

[2] F. Baader. Terminological cycles in a description logic with existential restrictions. LTCS-Report LTCS-02-02, Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology, Germany, 2002. See http://lat.inf.tu-dresden.de/research/reports.html.

[3] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. In *Proceedings of the Nineteenth International Joint Conference on Artificial Intelligence IJCAI-05*, Edinburgh, UK, 2005. Morgan-Kaufmann Publishers.

[4] F. Baader, R. Küsters, and R. Molitor. Computing least common subsumers in description logics with existential restrictions. In T. Dean, editor, *Proceedings of the 16th International Joint Conference on Artificial Intelligence (IJCAI'99)*, pages 96–101. Morgan Kaufmann, 1999.

[5] F. Baader, C. Lutz, and B. Suntisrivaraporn. Is tractable reasoning in extensions of the description logic $\mathcal{EL}$ useful in practice? In *Proceedings of the Methods for Modalities Workshop (M4M-05)*, Berlin, Germany, 2005.

[6] F. Baader, C. Lutz, and B. Suntisrivaraporn. CEL—a polynomial-time reasoner for life science ontologies. In U. Furbach and N. Shankar, editors, *Proceedings of the 3rd International Joint Conference on Automated Reasoning (IJCAR'06)*, volume 4130 of *Lecture Notes in Artificial Intelligence*, pages 287–291. Springer-Verlag, 2006.

[7] F. Baader, C. Lutz, and B. Suntisrivaraporn. Efficient reasoning in $\mathcal{EL}^+$. In *Proceedings of the 2006 International Workshop on Description Logics (DL2006)*, CEUR-WS, 2006. To appear.

[8] Franz Baader. Computing the least common subsumer in the description logic el w.r.t. terminological cycles with descriptive semantics. In Aldo

de Moor, Wilfried Lex, and Bernhard Ganter, editors, *ICCS*, volume 2746 of *Lecture Notes in Computer Science*, pages 117–130. Springer, 2003.

[9] Franz Baader. Least common subsumers and most specific concepts in a description logic with existential restrictions and terminological cycles. In Georg Gottlob and Toby Walsh, editors, *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 319–324. Morgan Kaufman, 2003.

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

[11] Franz Baader, Diego Calvanese, Deborah L. McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.

[12] Franz Baader, Carsten Lutz, and Boontawee Suntisrivaraporn. Is tractable reasoning in extensions of the description logic $\mathcal{EL}$ useful in practice? In *Journal of Logic, Language and Information, Special Issue on Method for Modality (M4M)*, 2007. To appear.

[13] Franz Baader, Rafael Peñaloza, and Boontawee Suntisrivaraporn. Pinpointing in the description logic $\mathcal{EL}^+$. In *Proceedings of the 30th German Conference on Artificial Intelligence (KI2007)*, LNAI, Osnabrück, Germany, 2007. Springer.

[14] Ron J. Brachman and James G. Schmolze. An Overview of the KL-ONE Knowledge Representation System. *Cognitive Science*, 9(2):171–216, 1985.

[15] S. Brandt. *Standard and Non-standard reasoning in Description Logics*. Ph.D. dissertation, Institute for Theoretical Computer Science, TU Dresden, Germany, 2006.

[16] S. Brandt and A.-Y. Turhan. Using non-standard inferences in description logics — what does it buy me? In *Proceedings of the KI-2001 Workshop on Applications of Description Logics (KIDLWS'01)*, number 44 in CEUR-WS, Vienna, Austria, September 2001. RWTH Aachen. Proceedings online available from http://SunSITE.Informatik.RWTH-Aachen.DE/Publications/CEUR-WS/Vol-44/.

[17] Sebastian Brandt. Polynomial time reasoning in a description logic with existential restrictions, GCI axioms, and—what else? In R. López de Mantáras and L. Saitta, editors, *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI-2004)*, pages 298–302. IOS Press, 2004.

[18] Sebastian Brandt and Jörg Model. Subsumption in w.r.t. hybrid tboxes. In Ulrich Furbach, editor, *KI*, volume 3698 of *Lecture Notes in Computer Science*, pages 34–48. Springer, 2005.

[19] Monika Rauch Henzinger, Thomas A. Henzinger, and Peter W. Kopke. Computing simulations on finite and infinite graphs. In *IEEE Symposium on Foundations of Computer Science*, pages 453–462, 1995.

[20] Martin Hofmann.

[21] J. Model. Subsumtion in el bezüglich hybrider tboxen. Diploma thesis, TU Dresden, Germany, 2005.

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

[23] N. Novaković. A subsumption reasoner in description logic $\mathcal{EL}$ w.r.t. hybrid tboxes. Student project report, Dresden University of Technology, Germany, 2007.

[24] Galen Ontology.
http://www.opengalen.org/.

[25] SNOMED CT$^{®}$ Ontology.
http://www.ihtsdo.org/our-standards/.

[26] Gene Ontology^{TM}.
http://www.geneontology.org.

[27] A. Rector and I. Horrocks. Experience building a large, re-usable medical ontology using a description logic with transitivity and concept inclusions, 1997.

[28] Ronald L. Rivest and Charles E. Leiserson. *Introduction to Algorithms*. McGraw-Hill, Inc., New York, NY, USA, 1990.

[29] K. Spackman. Spackman. managing clinical terminology hierarchies using algorithmic calculation. *J. of the American Medical Informatics Association*, Fall Symposium Special Issue, 2000.

[30] B. Suntisrivaraporn. Optimization and implementation of subsumption algorithms for the description logic $\mathcal{EL}$ with cyclic tboxes and general concept inclusion axioms. Master thesis, TU Dresden, Germany, 2005.

[31] Boontawee Suntisrivaraporn, Franz Baader, Stefan Schulz, and Kent Spackman. Replacing sep-triplets in snomed ct using tractable description logic operators. In Jim Hunter Riccardo Bellazzi, Ameen Abu-Hanna, editor, *Proceedings of the 11th Conference on Artificial Intelligence in Medicine (AIME'07)*, Lecture Notes in Computer Science. Springer-Verlag, 2007.

[32] A. Tarski. Lattice-theoretic fixpoint theorem and its applications. *Pacific Journal of Mathematics*, (5):285–309, 1955.