Technical University Dresden
Department of Computer Science

Master's Thesis on

Subsumption in the Description Logic $\mathcal{ELHI}f_{\mathcal{R}^+}$ w.r.t. General TBoxes

by

Quoc Huy Vu

born on 4^{th} December 1983 in Phu Tho, Vietnam

Advisors: Dipl.-Inf. Boontawee Suntisrivaraporn
and PD Dr. rer. nat. habil. Carsten Lutz

Overseeing Professor: Prof. Dr.-Ing. Franz Baader

Dresden, October 2008

Technische Universität Dresden

Author:	**Quoc Huy Vu**
Matrikel-Nr.:	**3368660**
Title:	**Subsumption in the Description Logic $\mathcal{ELHI}f_{\mathcal{R}^+}$ w.r.t. General TBoxes**
Degree:	**Master of Science**
Date of submission:	**02/10/2008**

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 I cited in the thesis.

Signature of Author

Abstract

Description Logics are a family of knowledge representation formalisms for representing and reasoning about conceptual knowledge. Every DL system has reasoning services as an important component that infer implicit knowledge from the one explicitly given. Standard reasoning problems include concept satisfiability, concept subsumption, ABox consistency and the instance problem. This work considers the concept subsumption service, which is considered to be the most "traditional" service.

Four years ago, a polynomial time algorithm for subsumption problem in the Description Logic $\mathcal{EL}$ was developed. After that, algorithms for different problems in tractable extensions of $\mathcal{EL}$ have been developing. These Description Logics are sufficient to represent many knowledge bases; however, there are ontologies requiring more expressive extensions of $\mathcal{EL}$. Specifically, GALEN, an important medical ontology, requires $\mathcal{ELHI}f_{\mathcal{R}^+}$, an intractable extension of $\mathcal{EL}$ that includes role hierarchies, inverse , functional and transitive roles. This motivates the extension of the polynomial time algorithm to $\mathcal{ELHI}f_{\mathcal{R}^+}$.

This thesis proposes two solutions for the subsumption problem in $\mathcal{ELHI}f_{\mathcal{R}^+}$. The first solution is to reduce $\mathcal{ELHI}f_{\mathcal{R}^+}$ to the less expressive DL $\mathcal{ELI}$, for which an algorithm is readily available. The second solution is to create an algorithm for the concept subsumption problem in $\mathcal{ELHI}f_{\mathcal{R}^+}$. This algorithm still runs in polynomial time in the simple case of $\mathcal{ELHI}f_{\mathcal{R}^+}$ that does not include inverse and functional roles.

Acknowledgements

First of all, I would like to express my deep gratitude and appreciation to Boontawee Suntisrivaraporn (Meng), my scientific advisor. Without the weekly meetings with him and getting his precious advice, I could not have finished this thesis in time.

I also wish to express my best thanks to Dr. Carsten Lutz, my co-advisor for his useful ideas and support, which have greatly influenced this work. My many thanks to him because he taught me how to start a research project and then write an academic report before this thesis.

I would like to thank Prof. Franz Baader, for giving a very motivating series of lectures, "Complexity and Logic", from which I have gained background knowledge for my thesis and passion for research. Also many thanks should go to the European Master's Program in Computational Logic and its director, Prof. Steffen Hölldobler, for their support and help during my study.

In addition, I am thankful to Dr. Khang Tran, my former supervisor at Hanoi University of Technology, and all my teachers at Free University of Bozen-Bolzano and Dresden University of Technolgy, who equipped me with the essential knowledge as well as the love for this field.

I would like to thank my dear secretaries Federica in Bolzano and Sylvia, Julia, Kerstin in Dresden for their help during my study.

To get the final writing of this work, I would like to express my best thank to Karen, Edward, Peter and Ilham, my most beloved of English teachers and friends who helped me to revise some sections of the final version.

Finally, to my parents, my little brother, Winnie and all my friends, thank you all for your eternal love and support for my whole life and this thesis.

Contents

Chapter 1

Introduction

Knowledge Representation (KR) is one of the most important fields being researched in artificial intelligence and cognitive science. Research in this field focuses on representing the world in a way that implicit knowledge can efficiently be found. There are some formalisms that are effective to represent knowledge, such as semantic networks and frames. Semantic networks, developed by M. R. Quillian [23], are graphs with labeled nodes and labeled edges. The nodes represent objects, concepts, and situations; while, the edges represent relationships between nodes. Frames were introduced by Minsky [21] to represent concepts and their properties. Each frame has a name, a list of direct super-frames and several slots, which are filled with values. Even though there are significant differences between semantic networks and frames, both are considered as network structures.

Since both semantic networks and frames lacked precision, formal semantics for them were essential. The semantics in these methods were represented by fragments of first-order logic, thus its reasoning service could be based on first-order logic provers. However, different fragments of first-order logic have different complexity and first-order logic provers are not always necessary. Thus, simpler and more specialized algorithms needed to be developed for these fragments. Based on these findings, a new method to define concepts through fragments of first-order logic was developed with the name "concept languages", which then changed to Description Logics (DLs).

Description Logics and Reasoning

DLs are a class of knowledge representation formalisms that represents the terminological knowledge of an application domain. The DL languages include sets of atomic concepts, roles, and constructors, which are formal and have logic-based semantics. The constructors are used to build complex concepts out of the atomic concepts and roles. In DL systems, DL languages play important roles both in DL knowledge bases and their associated reasoning services. A DL knowledge base consists of two components: a TBox, which defines the terminology of the application domain, and an ABox, which states facts about a specific "world".

In addition to the description language and the knowledge base, any DL system has its reasoning component, which derives implicit knowledge. Reasoning is not only a main task of DL systems but also the one that makes DLs distinguished from other KR formalisms. Reasoning services include: Concept satisfiability, concept subsumption, ABox consistency and the instance problem. The concept subsumption is the most "traditional" reasoning service and it is often supported by almost all DL systems. Given an ontology, the more general case of subsumption problem between concepts is the classification problem, where all the subsumption relations between concept names appearing in the ontology are listed. The complexity of algorithms for this problem is important in application. Almost all DL systems use intractable DLs, however, several years ago, applications using tractable DLs were also investigated.

The first intractability results for DLs were shown in the 1980s [11, 22]. Even with the simple DL $\mathcal{FL}_0$, which allows conjunction ($\sqcap$) and value restrictions ($\forall r.C$), when terminologies (TBoxes) are considered, the reasoning tasks become intractable. Specifically, with the simplest case of TBoxes (acyclic TBoxes), subsumption in $\mathcal{FL}_0$ and its extensions is coNP-hard [22]. The complexity of subsumption in $\mathcal{FL}_0$ w.r.t. cyclic TBoxes is PSpace-complete [1, 20]. [2] shows that it becomes ExpTime-complete in the presence of general concept inclusion axioms (GCIs), which are supported by all modern DL systems.

Since most DLs are intractable, as well as the need for expressive DLs supporting GCIs in applications, since the mid 1990s, DL researchers started investigating more and more expressive DLs, even though these DLs are worst-case intractable. Their goal is to find DLs which can be implemented using practical reasoning procedures. Their algorithms can be worst-case exponential or worse, but they behave well for practical applications [18]. This direction of research results in optimized DL systems using expressive DLs based on tableau algorithms [14, 16]. The most notable application of these DLs is the Web Ontology Language (OWL), which is recommended by World Wide Web Consortium (W3C) as an ontology language for the Semantic Web. Some problems in OWL can be solved by using reasoning in DLs. For example, the entailment problem in OWL can be reduced to concept satisfiability in $\mathcal{SHOIN}$ (OWL DL) and $\mathcal{SHIF}$ (OWL Lite) [17].

Even though there are a significant number of application using intractable DLs, tractable DLs are also valuable due to their efficiency in reasoning. Four years ago, a very interesting DL, named $\mathcal{EL}$, which allows conjunctions and existential restrictions, was found. The DL $\mathcal{EL}$ has better algorithmic properties than other DLs, notably having subsumption problem in $\mathcal{EL}$ w.r.t. general TBoxes solvable in polynomial time [3]. Furthermore, the classification of both cyclic and acyclic $\mathcal{EL}$-TBoxes is tractable [6]. This DL was then extended to other more expressive $\mathcal{EL}$-extensions, which are still tractable even w.r.t. GCIs, such as $\mathcal{EL}^{++}$. The DL $\mathcal{EL}^{++}$ allows role hierarchies, transitive roles and right-identity rules. It also includes bottom concept, which can be used to express disjointness of concept descriptions. The expressiveness of these tractable $\mathcal{EL}$-extensions is useful for ontology applications.

[9] has proposed a refined version of this polynomial time algorithm for implementation purposes, which gives better performance compared to optimized tableau-based DL systems. The algorithm is implemented for the DL $\mathcal{EL}^{+}$, which is simplified from $\mathcal{EL}^{++}$ by disallowing nominals and concrete domain. The reason for this simplification is that none of the considered ontologies uses nominals or concrete domain. Regarding experience with three bio-medical ontologies in this work, $\mathcal{EL}$-extensions are very useful in representing bio-medical ontologies.

Bio-medical ontologies

An ontology has a common definition as a "formal specification of how to represent the objects, concepts and other entities that are assumed to exist in some area of interest and the relationships that hold among them." The definition from the W3C is more concise: "An ontology defines the terms used to describe and represent an area of knowledge." The terms of a specific ontology are often restricted to the vocabularies used in that particular field or domain. Medical field is considered as one of the most important and beneficial fields that ontology can be applied to.

Gene Ontology (GO) [13], provides a controlled vocabulary to describe gene and gene product attributes in any organism. Terms in GO are biological vocabularies, which are structured so that you can query them at different levels. Currently, this ontology consists of thousands concept names and one transitive role "part-of".

SNOMED (Systematized Nomenclature of Medicine) [28] is a reference terminology for clinical terms. In 2002, the first version of SNOMED CT (SNOMED Clinical Terms), a terminological resource for clinical software applications, is released. SNOMED has contributed to many helpful clinical applications, such as individuals' health recording and retrieving systems.

GALEN (Generalised Architecture for Languages, Encyclopedias, and Nomenclatures in medicine) is a project supported by European Union to provide re-usable terminology resources for clinical systems [19]. GALEN technology has been developed and applied in theory and software tools for more than ten years. It provides not only re-used but also language-independent shared medical data. GALEN was first represented in its special DL language GALEN Representation and Integration Language, then it is translated to a normal Description Logic [15].

MOTIVATION AND SOLUTIONS

Tractable $\mathcal{EL}$ extensions are useful not only because of their polynomial property as mentioned above but also because their expressiveness is sufficient for several applications. For example, SNOMED uses $\mathcal{EL}$ with an acyclic TBox; Gene Ontology can be seen as an acyclic $\mathcal{EL}$ TBox with one transitive role. Even though large part of GALEN can be expressed in $\mathcal{EL}^+$, it requires additional expressivity that is no longer tractable.

In order to represent the full GALEN, we need to extend $\mathcal{EL}$ to include inverse, transitive, functional roles and role hierarchy, named $\mathcal{ELHI}f_{\mathcal{R}^+}$. This DL is intractable because its sub-language $\mathcal{ELI}$ has been proved intractable[7]. At this point, it is natural to think about an algorithm for the DL $\mathcal{ELHI}f_{\mathcal{R}^+}$. This algorithm can be developed from the polynomial one for the DL $\mathcal{EL}$.

In spite of successes of this polynomial time algorithm in the DL $\mathcal{EL}$ and its tractable extensions, to the best of the author's knowledge, no one has investigated the link to other intractable extensions of $\mathcal{EL}$. In the work undertaken on the project before this thesis [29], the algorithm in [3] has been improved to be able to work with $\mathcal{ELI}$ w.r.t. general TBoxes. The algorithm in [29] requires exponential time in the worst case and it is optimal since it is shown in [7] that the subsumption problem is ExpTime-complete. The thesis proposes two directions to solve the subsumption problem in $\mathcal{ELHI}f_{\mathcal{R}^+}$ and its fragments.

The first direction is to reduce the input general TBox to an $\mathcal{ELI}$ general TBox. This way is very convenient for $\mathcal{ELI}f$ general TBox because it can be reduced to an $\mathcal{ELI}$ general TBox, which has the size linear in the size of the input $\mathcal{ELI}f$ general TBox. For $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes we need to reduce out transitive roles, functional roles and then role hierarchies step by step to get $\mathcal{ELI}$ general TBoxes. The size of the input TBox polynomially increases after each step of the reduction. When having the new general $\mathcal{ELI}$ general TBoxes, the algorithm in [29] is applied.

The second direction is to devise a direct algorithm for subsumption problem in $\mathcal{ELHI}f_{\mathcal{R}^+}$, which is extended from the polynomial algorithm in [3], for these intractable $\mathcal{EL}$-extensions. Similar to the polynomial one, we build a mapping to represent the subsumption relations between concept descriptions. The mapping is a *labeling function* of a *completion graph.* The most notable difference of this algorithm from the algorithm in [3] is that its set of nodes grows during rule applications, whereas the nodes in [3] are fixed. Although the proposed algorithm is an exponential time algorithm, it runs in polynomial time in the special case of $\mathcal{ELH}_{\mathcal{R}^+}$ with GCIs. This subsumption algorithm not only computes subsumption between two given concept names in the input TBox $\mathcal{T}$; but also *classifies* $\mathcal{T}$, i.e., it simultaneously computes the subsumption relationships between *all* pairs of concept names occurring in $\mathcal{T}$.

After this introduction, the remainder of the thesis is organized as follows. Chapter 2 first introduces DL $\mathcal{EL}$ and then it describes the basic concept about an important tractable extension of $\mathcal{EL}$, named $\mathcal{EL}^{++}$. After that, this chapter presents the intractable DL $\mathcal{ELHI}f_{\mathcal{R}^+}$ and its sub-languages, which will be considered during the remainder of the paper. Chapter 2 ends by giving a normal form for the DL $\mathcal{ELHI}f_{\mathcal{R}^+}$.

Chapter 3 presents GALEN projects and GALEN ontology, which can be formulated using DL $\mathcal{ELHI}f_{\mathcal{R}^+}$. The basic concept of GALEN and GRAIL, the developing language for GALEN, is presented in the first section of this chapter. The second section shows the method of representing GRAIL concept using DL $\mathcal{ELHI}f_{\mathcal{R}^+}$. After that, Chapters 4 and 5 present two directions to solve the subsumption problem in $\mathcal{ELHI}f_{\mathcal{R}^+}$ and its sub-languages.

In Chapter 4, the first solution is presented: reducing the input general TBoxes to $\mathcal{ELI}$ general TBoxes. The chapter begins by a special case of $\mathcal{ELHI}f_{\mathcal{R}^+}$, named $\mathcal{ELI}f$. In this case, after the reduction from $\mathcal{ELI}f$-TBoxes to $\mathcal{ELI}$ general TBoxes, the new TBoxes are linear in the sizes of the input ones. After that, other reductions are presented to reduce an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox to an $\mathcal{ELI}$ general TBox.

Chapter 5 proposes direct algorithms for subsumption problems in intractable extensions of $\mathcal{EL}$ and the proof of their termination, soundness and completeness. The last section of this chapter proves an interesting property that the algorithm works polynomially in the case of DL $\mathcal{ELH}_{\mathcal{R}^+}$ with GCIs.

The last chapter, the conclusion, summarizes the contributions of the present thesis and sheds light on some of the next essential steps, such as optimization and implementation of the algorithms.

Chapter 2

Description Logics

In this chapter, the DL $\mathcal{EL}$ and its extensions are explored. After the DL $\mathcal{EL}$ is introduced, a notable tractable $\mathcal{EL}$ extension, named $\mathcal{EL}^{++}$, is given together with its reasoning service. Then this chapter presents the DL $\mathcal{ELHI}f_{\mathcal{R}^+}$, an intractable $\mathcal{EL}$ extension that includes role hierarchies, inverse, transitive and functional roles. Finally, a normal form of the $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes is introduced.

2.1 The Description Logic $\mathcal{EL}$

The syntax and semantics of the basic DL $\mathcal{EL}$ are defined bellow.

Name	Syntax	Semantics
top	$\top$	$\Delta^{\mathcal{I}}$
conjunction	$C \sqcap D$	$C^{\mathcal{I}} \cap D^{\mathcal{I}}$
existential restriction	$\exists r.C$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y \in \Delta^{\mathcal{I}} : (x, y) \in r^{\mathcal{I}} \wedge y \in C^{\mathcal{I}}\}$

Table 2.1: Syntax and Semantics of $\mathcal{EL}$

Definition 2.1.1. The syntax of $\mathcal{EL}$-concept descriptions is inductively defined as follows:

- All concept names are $\mathcal{EL}$-concept descriptions;
- if C, D are two $\mathcal{EL}$-concept descriptions, and r is a role, then $\top$, $C \sqcap D$, and $\exists r.C$ are $\mathcal{EL}$-concept descriptions.

Definition 2.1.2. The semantics of $\mathcal{EL}$-concept descriptions is defined in terms of an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. The *domain* $\Delta^{\mathcal{I}}$ is a non-empty set of individuals and the *interpretation function* $\cdot^{\mathcal{I}}$ maps each concept name $A \in N_{con}$ to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ and each role name $r \in N_{role}$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The function $\cdot^{\mathcal{I}}$ can be extended to arbitrary concept descriptions as shown in the third column of Table 2.1.

Though lightweight, the DL $\mathcal{EL}$ is sufficient to express notions in the clinical domain. Some examples from GALEN ontology are given below:

$\mathsf{Person} \sqcap \exists\ \mathsf{playsSocialRole.DoctorRole}$
represents the concept of a Doctor, and

$\exists\ \mathsf{hasClinicalSpeciality.(ClinicalSpeciality} \sqcap (\exists\ \mathsf{hasState.Surgery}))$
is a $\mathsf{Surgeon}$.

2.2 A tractable extension of $\mathcal{EL}$: $\mathcal{EL}^{++}$

Syntax and semantics of DL $\mathcal{EL}^{++}$ are defined in Table 2.2 in combination with Table 2.1

Name	Syntax	Semantics
bottom	$\bot$	$\emptyset$
nominal	$\{a\}$	$\{a^{\mathcal{I}}\}$
concrete domain	$p(f_1, \ldots, f_k)$ for $p \in \mathcal{P}^{\mathcal{D}_j}$	$\{x \in \Delta^{\mathcal{I}} \mid \exists y_1, \ldots, y_k \in \Delta^{\mathcal{D}_j} : f_i^{\mathcal{I}}(x) = y_i$ for $1 \leq i \leq k \wedge (y_1, \ldots, y_k) \in \mathcal{P}^{\mathcal{D}_j}\}$
GCI	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
RI	$r_1 \circ \cdots \circ r_k \sqsubseteq r$	$r_1^{\mathcal{I}} \circ \cdots \circ r_k^{\mathcal{I}} \subseteq r^{\mathcal{I}}$
concept assertion	$C(a)$	$a^{\mathcal{I}} \in C^{\mathcal{I}}$
role assertion	$r(a, b)$	$(a^{\mathcal{I}}, b^{\mathcal{I}}) \in r^{\mathcal{I}}$

Table 2.2: Extended Syntax and Semantics of $\mathcal{EL}^{++}$

Definition 2.2.1. The syntax of $\mathcal{EL}^{++}$-concept descriptions are formed using the constructors shown in Table 2.1 and the upper part of Table 2.2.

The DL $\mathcal{EL}^{++}$ can include a set of concrete domains $\{\mathcal{D}_1, \ldots, \mathcal{D}_n\}$, in which an arbitrary element $\mathcal{D}_j$ is defined as follows.

Definition 2.2.2. (*concrete domain.*)
A *concrete domain* $\mathcal{D}_j$ is a pair $(\Delta^{\mathcal{D}_j}, \mathcal{P}^{\mathcal{D}_j})$ such that $\Delta^{\mathcal{D}_j}$ is a set and $\mathcal{P}^{\mathcal{D}_j}$ is

a set of *predicate names*. Each $p \in \mathcal{P}^{\mathcal{D}_j}$ is associated with an extension $p^{\mathcal{D}_j} \in (\Delta^{\mathcal{D}_j})^n$ and is linked to the DL through a set of *feature names* $\{f_1, \ldots, f_k\}$.

Definition 2.2.3. The semantics of an $\mathcal{EL}^{++}$-concept description is defined in terms of an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. The *domain* $\Delta^{\mathcal{I}}$ is a non-empty set of individuals and the *interpretation function* $\cdot^{\mathcal{I}}$ maps each concept name $A \in N_{con}$ to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ and each role name $r \in N_{role}$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. The function $\cdot^{\mathcal{I}}$ can be extended to arbitrary concept descriptions as shown in the third column of Table 2.1 and Table 2.2.

Definition 2.2.4. $\mathcal{EL}^{++}$ Constraint Box (CBox)
An $\mathcal{EL}^{++}$-CBox $\mathcal{T}$ is defined as a finite set of *general concept inclusions* and *role inclusions*, which are defined below:

- Let C, D be $\mathcal{EL}^{++}$-concept descriptions. Then $C \sqsubseteq D$ is called a *general concept inclusion (GCI)*.
- Let $r_1 \circ \cdots \circ r_k$ and r be $\mathcal{EL}^{++}$-role names. Then $r_1 \circ \cdots \circ r_k \subseteq r$ is called a *role inclusion (RI)*.

Definition 2.2.5. An interpretation $\mathcal{I}$ *satisfies* a GCI $C \sqsubseteq D$ or a RI $r_1 \circ \cdots \circ r_k \subseteq r$, if its semantics in the third column of Table 2.2 is satisfied. We notice that in the definition of RI's semantics, $\circ$ denotes composition of binary relations. An interpretation $\mathcal{I}$ is a *model* of the CBox $\mathcal{T}$ if $\mathcal{I}$ satisfies all general concept inclusions and role inclusions in $\mathcal{T}$.

To represent medical ontologies we often use a subclass of $\mathcal{EL}^{++}$, called $\mathcal{EL}^{+}$ [9]. In DL $\mathcal{EL}^{+}$, some features in $\mathcal{EL}^{++}$ are dropped out, they are nominal and concrete domain. This DL has been implemented and seen the performing advantage in CEL [4].

However, there are medical ontologies that need other features. In particular, GALEN ontology requires an extension of $\mathcal{EL}$ having role hierarchies, and inverse roles, functional and transitive roles.

2.3 An intractable extension of $\mathcal{EL}$: $\mathcal{ELHI}f_{\mathcal{R}^+}$

In the previous section, $\mathcal{EL}^{++}$ has been shown as one of the most useful extensions of $\mathcal{EL}$ that is still tractable. Even though this DL is sufficient for many medical ontologies, there are demands from some ontologies, such as GALEN, for a more expressive DL. This section considers the DL $\mathcal{ELHI}f_{\mathcal{R}^+}$, which is logical representation formalism for GALEN.

The DL $\mathcal{ELHI}f_{\mathcal{R}^+}$ is extended from $\mathcal{EL}$ with role hierarchies, inverse, functional and transitive roles. The syntax and semantics of additional expressivity is shown in Table 2.3.

Definition 2.3.1. The syntax of $\mathcal{ELHI}f_{\mathcal{R}^+}$-concept descriptions are inductively defined as follows:

Name	Syntax	Semantics
inverse role	r^-	$\{(d,c) \mid (c,d) \in r^{\mathcal{I}}\}$
transitive axiom	$r \circ r \sqsubseteq r$	$r^{\mathcal{I}} \circ r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$
functional role	$\top \sqsubseteq (\leq 1r)$	$\lvert\{e \mid (d,e) \in r^{\mathcal{I}}\}\rvert \leq 1$ for all $d \in \Delta^{\mathcal{I}}$
GCI	$C \sqsubseteq D$	$C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$
role hierarchy	$r \sqsubseteq s$	$r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$

Table 2.3: Extended Syntax and Semantics of $\mathcal{ELHI}f_{\mathcal{R}^+}$

- All elements in the *set of concept names* N_{con} are concept descriptions;
- if C, D are two concept descriptions, and r is a role in the *set of roles* N_{role}, then $\top$, $C \sqcap D$, $\exists r.C$, and $\exists r^-.C$ are concept descriptions.

Definition 2.3.2. The semantics of an $\mathcal{ELHI}f_{\mathcal{R}^+}$-concept description is defined in terms of an *interpretation* $\mathcal{I} = (\Delta^{\mathcal{I}}, \cdot^{\mathcal{I}})$. The *interpretation domain* $\Delta^{\mathcal{I}}$ is a non-empty set of individuals and the *interpretation function* $\cdot^{\mathcal{I}}$ maps each concept name $A \in N_{con}$ to a subset $A^{\mathcal{I}}$ of $\Delta^{\mathcal{I}}$ and each role name $r \in N_{role}$ to a binary relation $r^{\mathcal{I}} \subseteq \Delta^{\mathcal{I}} \times \Delta^{\mathcal{I}}$. For arbitrary concept descriptions, the semantics of $\cdot^{\mathcal{I}}$ is defined as it is in the third columns of Table 2.1 and Table 2.3.

From the definition of inverse roles, we notice that $(r^-)^- = r$ for all role $r \in N_{role}$. Therefore from now, when we write r, it can be either a role or an inverse role.

Here we consider a complex concept description appeared in the GALEN ontology

SmoothMuscleContractionProcess $\sqcap\, \exists hasProcessActivity.$(ProcessActivity $\sqcap$ ($\exists hasState.$ inactive)) $\sqcap\, \exists hasFunction^-.$ MuscleOfUrinaryBladder

This is the definition of UrinaryBladderAtonia.

Now we define general TBoxes in the DL $\mathcal{ELHI}f_{\mathcal{R}^+}$.

Definition 2.3.3. ($\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox). An $\mathcal{ELHI}f_{\mathcal{R}^+}$ *general TBox* $\mathcal{T}$ is defined as a finite set of GCIs, which is defined before, and functional axioms, role hierarchies, transitive axioms defined below.

- Let r be an $\mathcal{ELHI}f_{\mathcal{R}^+}$-role. Then $\top \sqsubseteq (\leq 1r)$ is called a *functional axiom.*

- Let r, s be $\mathcal{ELHI}f_{\mathcal{R}^+}$-roles. Then $r \sqsubseteq s$ is called a *role hierarchy.*
- Let r be an $\mathcal{ELHI}f_{\mathcal{R}^+}$-role. Then $r \circ r \sqsubseteq r$ is called a *transitive axiom* and $\top \sqsubseteq (\leq 1r)$ is called a *functional axiom.*

Definition 2.3.4. An interpretation $\mathcal{I}$ *satisfies* a GCI, a functional axiom, a role hierarchy axiom or a transitive axiom iff its semantics in the third column of Table 2.3 is satisfied. An interpretation $\mathcal{I}$ is a *model* of the $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox $\mathcal{T}$ if $\mathcal{I}$ satisfies all axioms appeared in $\mathcal{T}$.

Motivated by GALEN and an example in [4], we introduce an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox, named $\mathcal{T}$, as in Figure 2.1. There are GCIs, role hierarchies,

$$
\begin{array}{rcl}
\mathsf{Endocardium} & \sqsubseteq & \mathsf{Tissue} \sqcap \exists \mathsf{have}^-.\mathsf{HeartWall} \sqcap \exists \mathsf{have}^-.\mathsf{HeartValve} \\
\mathsf{HeartWall} & \sqsubseteq & \mathsf{BodyWall} \sqcap \exists \mathsf{part\text{-}of}.\mathsf{Heart} \\
\mathsf{HeartValve} & \sqsubseteq & \mathsf{BodyValve} \sqcap \exists \mathsf{part\text{-}of}.\mathsf{Heart} \\
\mathsf{Endocarditis} & \sqsubseteq & \mathsf{Inflammation} \sqcap \exists \mathsf{has\text{-}loc}.\mathsf{Endocardium} \\
\mathsf{Inflammation} & \sqsubseteq & \mathsf{Disease} \sqcap \exists \mathsf{has\text{-}loc}.\mathsf{Tissue} \\
\mathsf{HeartDisease} & \equiv & \mathsf{Disease} \sqcap \exists \mathsf{has\text{-}loc}.\mathsf{Heart} \\
\mathsf{have}^- & \sqsubseteq & \mathsf{has\text{-}loc} \\
\mathsf{part\text{-}of}^- & \sqsubseteq & \mathsf{have} \\
\mathsf{has\text{-}loc} \circ \mathsf{has\text{-}loc} & \sqsubseteq & \mathsf{has\text{-}loc} \\
\mathsf{Person} & \sqsubseteq & \exists \mathsf{hasUnique}.\mathsf{Heart} \\
\top & \sqsubseteq & (\leq 1 \mathsf{hasUnique})
\end{array}
$$

Figure 2.1: An $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox

inverse, functional and transitive roles in this TBox.

Given an $\mathcal{ELHI}f_{\mathcal{R}^+}$-TBox $\mathcal{T}$, the most relevant inference problems are defined as follows.

- *Concept satisfiability.* A concept description C is *satisfiable* w.r.t. $\mathcal{T}$ if there exists a model $\mathcal{I}$ of $\mathcal{T}$ such that $C^{\mathcal{I}} \neq \emptyset$.
- *Concept subsumption.* Given two concept descriptions C and D, C is *subsumed* by D w.r.t. $\mathcal{T}$, denoted by $\mathcal{T} \models C \sqsubseteq D$ or $C \sqsubseteq_{\mathcal{T}} D$ if and only if $C^{\mathcal{I}} \subseteq D^{\mathcal{I}}$ for all models $\mathcal{I}$ of $\mathcal{T}$.
- *ABox consistency.* An ABox $\mathcal{A}$ is *consistent* w.r.t. $\mathcal{T}$ if $\mathcal{A}$ and $\mathcal{T}$ have a common model.
- *The instance problem.* An individual name a is an instance of a concept C in an ABox $\mathcal{A}$ w.r.t. $\mathcal{T}$ if $a^{\mathcal{I}} \in C^{\mathcal{I}}$ for every common model $\mathcal{I}$ of $\mathcal{A}$ and $\mathcal{T}$.

In this thesis, the subsumption algorithms for different DLs are considered. There are some reasons to concentrate on this reasoning task. The first one is that subsumption is the most "traditional" reasoning service in DLs, as written in [5]. The second reason is that, as mentioned in the introduction, our subsumption algorithm has the ability to simultaneously compute the subsumption relationships between all pairs of concept names in the input general TBoxes.

With respect to the $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox in Figure 2.1, $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \mathsf{HeartDisease}$ is an example of concept subsumption. This subsumption relation can be reasoned as follows: Since $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \mathsf{Inflammation}$ and $\mathsf{Inflammation} \sqsubseteq_{\mathcal{T}} \mathsf{Disease}$ we have $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \mathsf{Disease}$.

More over, $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \exists \mathsf{has\text{-}loc.Endocardium}$ and $\mathsf{Endocardium} \sqsubseteq_{\mathcal{T}} \mathsf{have}^{-}.\mathsf{HeartWall}$, thus, $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \exists \mathsf{has\text{-}loc.have}^{-}.\mathsf{HeartWall}$. Besides, $\mathsf{have}^{-} \sqsubseteq_{\mathcal{T}} \mathsf{has\text{-}loc}$, we have $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \exists \mathsf{has\text{-}loc.has\text{-}loc.HeartWall}$. Since $\mathsf{has\text{-}loc} \circ \mathsf{has\text{-}loc} \sqsubseteq_{\mathcal{T}} \mathsf{has\text{-}loc}$, $\mathsf{part\text{-}of} \sqsubseteq_{\mathcal{T}} \mathsf{has\text{-}loc}$ and $\mathsf{HeartWall} \sqsubseteq_{\mathcal{T}} \exists \mathsf{part\text{-}of.Heart}$, we get $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \exists \mathsf{has\text{-}loc.Heart}$.

Therefore, $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \mathsf{Disease} \sqcap \exists \mathsf{has\text{-}loc.Heart}$, which means that $\mathsf{Endocarditis} \sqsubseteq_{\mathcal{T}} \mathsf{HeartDisease}$.

Theorem 3 in [7] states the complexity of subsumption problem in $\mathcal{ELI}$ w.r.t. general TBoxes, which is repeated in the following theorem.

Theorem 2.3.5. *The subsumption problem in $\mathcal{ELI}$ w.r.t. general TBoxes is* ExpTime*-complete.*

The next chapter will give us a polynomial reduction from $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes to $\mathcal{ELI}$ general TBoxes, thus the subsumption problem in the DL $\mathcal{ELHI}f_{\mathcal{R}^+}$ w.r.t. general TBoxes is ExpTime-complete. We conclude this section by giving a theorem about the complexity of subsumption problem in $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes.

Theorem 2.3.6. *The subsumption problem in $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes is* ExpTime*-complete.*

In the next section, a normal form for an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox is defined to give convenience to the subsumption algorithms.

2.4 A normal form for $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes

Given an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox $\mathcal{T}$ having the set of concept names N_{con}, we use $N_{\mathcal{T}}$ to denote the set of *basic concept descriptions* for $\mathcal{T}$:

$$N_{\mathcal{T}} := N_{con} \cup \{\top\}$$

Now, a normal form for $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes can be defined as follows.

Definition 2.4.1. (Normal Form for $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes)
An $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox $\mathcal{T}$ is in *normal form* if all concept inclusions in $\mathcal{T}$ have one of the following forms

$$\begin{aligned} A &\sqsubseteq B \\ A_1 \sqcap A_2 &\sqsubseteq B \\ A &\sqsubseteq \exists r.B \\ \exists r.A &\sqsubseteq B \end{aligned}$$

where $A, A_1, A_2, B \in N_{\mathcal{T}}$, and r is a role or inverse role.

Using the method presented in [2], new concept names are introduced to turn TBox $\mathcal{T}$ into a normalized TBox $\mathcal{T}'$. The new TBox $\mathcal{T}'$ is a *conservative extension* of $\mathcal{T}$, i.e., every model of $\mathcal{T}'$ is a model of $\mathcal{T}$ and every model of $\mathcal{T}$ can be extended to a model of $\mathcal{T}'$.

It can be proved that this transformation can be done in linear time, yielding a normalized TBox $\mathcal{T}'$ whose size is *linear* in the size of $\mathcal{T}$. The *size* $|\mathcal{T}|$ of a TBox $\mathcal{T}$ is defined as the number of symbols necessary to write down $\mathcal{T}$.

Lemma 2.4.2. *Subsumption w.r.t. $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes can be reduced in linear time to subsumption w.r.t. normalized $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes.*

Proof. Let $\mathcal{T}$ be an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox. For every $\mathcal{ELHI}f_{\mathcal{R}^+}$ concept description, the normalization rules are defined modulo commutativity of conjunction as in Table 2.4. The $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox is converted

NF1	$C \sqcap \hat{D} \sqsubseteq E$	$\longrightarrow$	$\{\hat{D} \sqsubseteq A, C \sqcap A \sqsubseteq E\}$
NF2	$\exists r.\hat{C} \sqsubseteq D$	$\longrightarrow$	$\{\hat{C} \sqsubseteq A, \exists r.A \sqsubseteq D\}$
NF3	$\hat{C} \sqsubseteq \hat{D}$	$\longrightarrow$	$\{\hat{C} \sqsubseteq A, A \sqsubseteq \hat{D}\}$
NF4	$B \sqsubseteq \exists r.\hat{C}$	$\longrightarrow$	$\{B \sqsubseteq \exists r.A, A \sqsubseteq \hat{C}\}$
NF5	$B \sqsubseteq C \sqcap D$	$\longrightarrow$	$\{B \sqsubseteq C, B \sqsubseteq D\}$

where $C, D, E, \hat{C}, \hat{D}$ are concept descriptions over $N_{\mathcal{T}}$, N_{role} such that $\hat{C}, \hat{D} \notin N_{\mathcal{T}}$; and A is a new concept name.

Table 2.4: Normalization Rules

into normal form using that set of rules in two phases:

1. exhaustively apply rules **NF1** and **NF2**;

2. exhaustively apply rules **NF3** to **NF5**.

The number of rule applications in Phase 1 is linear in the size of $\mathcal{T}$. Each time one of the Rules **NF1**, **NF2** is applied, the size of $\mathcal{T}$ increases only by a constant. Hence, the size of the resulting TBox after Phase 1 is linear in the size of the original one.

After finishing Phase 1, all concept inclusions have the normal forms and two forms $A \sqsubseteq D$, $\exists r.A \sqsubseteq D$, where $A \in N_{\mathcal{T}}$, D is a concept description and r is a role. Therefore, Rule **NF5** in Phase 2 cannot make a quadratic blowup due to the duplication of the concept B. The application of Rule **NF3** or Rule **NF4** only increases the size of $\mathcal{T}$ by a constant. Thus the size of the resulting general TBox is still linear in the size of the original one. It is easy to see that all concept inclusions that are not in normal form after Phase 1 are normalized in Phase 2.

Therefore, applying the rules exhaustively produces a normalized TBox which is linear in the size of the input $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox. Since each rule application linearly increases the size of the $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox, the process of applying those rules runs in linear steps.

Besides, it is readily checked that each rule application takes linear time. Therefore the normalization is linear.

Chapter 3

The GALEN Ontology

3.1 The GALEN Project and GRAIL Language

GALEN, a European Union funded project, has been developed to represent clinical information in a new way. The project builds an ontology, the Common Reference Model, to represent medical concepts independently of any application [25]. Many clinical systems, such as electronic health care records (EHCRs), decision support systems and computer-based multilingual coding systems for medicine, benefit from this ontology.

The GALEN Programme represents the overall development of the technology, which has included several research projects, including Framework III (GALEN project) and Framework IV (GALEN-IN-USE project). In the early stage, the GALEN Programme mainly constructed a concept model language, named GALEN Representation and Integration Language (GRAIL). At the same time, different structures of GALEN Common Reference Model were experimented.

In later stages, during the late 1990s, researchers concentrated on implementation of GRAIL and the terminology server, as well as the development of the GALEN Common Reference Model. In addition, they developed tools and techniques working with GALEN Common Reference Model. The development of tools and techniques helped to develop and maintain the model better.

In the second phase, the purpose of GALEN-IN-USE project is to develop terminologies for medical procedures and to demonstrate the GALEN technology in data entry and natural language modules for commercial clinical systems. The ontology has been used for building the architecture and infrastructure of clinical information systems in the EU-funded SynEx Project. The GALEN ontology is also used in the UK's Prodigy Drug Prescribing Project.

The GALEN Programme is not totally funded by European Union now. The members of the GALEN Programme founded a non-profit organization named OpenGALEN to expand its results and to find other related technologies for the GALEN ontology.

GRAIL is a concept modeling language developed in the GALEN Programme to afford the demand of representation for GALEN [26, 27]. Due to the main purpose of GRAIL, it includes transitive roles and role hierarchies [24]. However, in comparison with other DLs, it lacks a number of common properties, such as cardinality, negation, disjunction, value restrictions. GRAIL's special features and unusual syntax make it more accessible to domain experts. The main forms of GRAIL statements are summarized below:

- C `which` $\langle r_1 C_1 \ldots r_n C_n \rangle$

  The GRAIL `which` statement is used to form concept terms based on concepts and existential restrictions.

  For example,

  Person `which` isOwnerOf RoadVehicle
  represents the concept of a person who owns a road vehicle (not a train or a boat).

- C `newSub` CN

  The GRAIL `newSub` statement is used to state subsumption between concepts.

  For instance, a car is a kind of RoadVehicle can be represented as

  RoadVehicle `newSub` Car

- C `name` CN

  There are complicated concepts created by `which` and other features, so it is necessary for GRAIL to use `name` as an aliasing mechanism, which is equivalent to concept definition. For example,

  Car `which` hasOwner ((Person `which` hasAge old) `which` hasSex male).

  then the concept ((Person `which` hasAge old)`which` hasSex male) should be named as OldMan to be simple as follows:

  ((Person `which` hasSex male) `which` hasAge old) `name` OldMan.

- C `topicNecessarily` $\langle r_1 C_1 \ldots r_n C_n \rangle$

  The GRAIL `topicNecessary` statement is used for situations where a criterion is mandatory but not part of the definition.

  For example,

  (Person `which` isOwnerOf RoadVehicle) `topicNecessarily` hasAge old
  gives the necessary condition for the owner of a road vehicle that he or she must be old.

- r `newAttribute` r_1 r_2 k

  GRAIL statement `newAttribute` is used to create attributes. In the above form, r is an existing role, r_1 and r_2 are a pair of new roles, k is a keyword which determines if these new roles are roles or attributes. There are four possible values for k: `oneOne` (both r_1 and r_2 are attributes), `oneMany` (r_1 is an attribute), `manyOne` (r_2 is an attribute), `manyMany` (both r_1 and r_2 are not attributes).

  For example,

  DomainAttribute `newAttribute` hasOwner isOwnerOf allAll manyOne
  This introduces a new attribute hasOwner, which has an inverse isOwnerOf. The keyword manyOne means that one can own many objects but an object belongs to only one owner.

- RN `transitiveDown`

  Transitive relations are necessary for a medical ontology, for example,

  hasPart `transitiveDown`
  states that, if we have

  Vehicle `hasPart` Wheel

  Wheel `hasPart` Tyre
  then we have, Vehicle `hasPart` Tyre

- r_1 `addSub` r_2

  `addSub` is used to add an additional role inclusion into the whole role hierarchy.

  For example,

  leftSided `addSub` bothSided
  gives us that if a car has blue color on both sides, then it has blue color on its left side.

  There is another GRAIL statement, which is similar to `addSub`, named `addSuper`.

  For example,

  bothSided `addSuper` leftSided
  gives us the same semantic as the above example with `addSub`.

3.2 Translating GRAIL into $\mathcal{ELHI}f_{\mathcal{R}^+}$

The syntax and semantics of $\mathcal{ELHI}f_{\mathcal{R}^+}$ is appropriate for GRAIL. In his PhD thesis [15], Horrocks has presented the translation from GRAIL into $\mathcal{ALCH}f_{\mathcal{R}^+}$. The DL $\mathcal{ELHI}f_{\mathcal{R}^+}$ not only supports all GRAIL's statements that DL $\mathcal{ALCH}f_{\mathcal{R}^+}$ supports but also supports inverse roles that are not explicitly included in $\mathcal{ALCH}f_{\mathcal{R}^+}$. Moreover, $\mathcal{ELHI}f_{\mathcal{R}^+}$ does not include those extra Boolean constructors that $\mathcal{ALCH}f_{\mathcal{R}^+}$ offers but GRAIL lacks.

Following the approach in [15], we here present the translation from GRAIL into $\mathcal{ELHI}f_{\mathcal{R}^+}$ as follows.

Regarding the forms of GRAIL statements shown in the previous section, GRAIL concept statements have the equivalent $\mathcal{ELHI}f_{\mathcal{R}^+}$ concept descriptions and axioms as in the right column of Table 3.1.

The `which` statement is the combination of existential restrictions and conjunctions.

The `newSub` and `name` are axioms used to define concept.

The last concept statement, `topicNecessarily` statement's purpose is to add GCI axioms to the knowledge base.

GRAIL	$\mathcal{ELIHF}_{\mathcal{R}^+}$
C `which` $\langle r_1 C_1 \ldots r_n C_n \rangle$	$C \sqcap \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_n.C_n$
C `newSub` CN	$CN \sqsubseteq C$
C `name` CN	$CN \doteq C$
C `topicNecessarily` $\langle r_1 C_1 \ldots r_n C_n \rangle$	$C \sqsubseteq \exists r_1.C_1 \sqcap \ldots \sqcap \exists r_n.C_n$

Table 3.1: GRAIL concept statements and equivalent $\mathcal{ELHI}f_{\mathcal{R}^+}$

The translation of GRAIL concept statements to equivalent $\mathcal{ELHI}f_{\mathcal{R}^+}$-concept descriptions has been described, GRAIL role statements are considered below. The `newAttribute` statement is used to introduce pairs of

GRAIL	$\mathcal{ELHI}f_{\mathcal{R}^+}$
r `newAttribute` r_1 r_2 k	$r_1 \sqsubseteq r$, $r_2 \doteq r_1^-$
r `transitiveDown`	$r \circ r \sqsubseteq r$
r_1 `addSub` r_2	$r_2 \sqsubseteq r_1$

Table 3.2: GRAIL role axioms and equivalent $\mathcal{ELHI}f_{\mathcal{R}^+}$

primitive roles. The general form of `newAttribute` statements, as mentioned above, is

r `newAttribute` r_1 r_2 k

We have r_1 is a new role that satisfies $r_1 \sqsubseteq r$. In addition, the new role r_2 is the inverse role of r_1, i.e., $r_2 \doteq r_1^-$. Thus the translation for `newAttribute` is stated as in the first line of Table 3.2.

Since `newAttribute` statement allows roles to be declared functional through the keyword k, the functional axioms in $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes are also necessary.

The `transitiveDown` statement is directly equivalent to a transitive axiom in DL $\mathcal{ELHI}f_{\mathcal{R}^+}$.

The `addSub` statement, which is used to build role hierarchy, is translated to a role hierarchy in $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes as in Table 3.2.

Chapter 4

Reductions of $\mathcal{ELHI}f_{\mathcal{R}^+}$ General TBoxes

4.1 Reduction from $\mathcal{ELHI}f_{\mathcal{R}^+}$ to $\mathcal{ELHI}f$

Since roles that are both transitive and functional are not clearly useful in practice, we assume that functionality can be asserted on *simple roles*, which are defined below.

Definition 4.1.1. A role that is neither a transitive role nor a super role of a transitive one is a *simple role*.

In order to eliminate all the transitive axioms $r \circ r \sqsubseteq r$ in the input $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox $\mathcal{T}$, a set of reduction rules in Table 4.1 is applied as follows. Firstly, GCIs in the right column of Table 4.1 are added if the conditions in the left column are satisfied. In these new GCIs, X is a new concept name. This process terminates when all possible set of three rules have been considered. Finally, all transitive axioms are eliminated and a new TBox $\mathcal{T}'$ without transitive axioms is obtained.

Rule	Original GCI and RI	New GCIs
RR	$r \circ r \sqsubseteq r$	$\exists r.A \sqsubseteq X$
	$\exists s.A \sqsubseteq B$	$X \sqsubseteq B$
	$r \sqsubseteq s$	$\exists r.X \sqsubseteq X$

Table 4.1: Reduction rule from $\mathcal{ELHI}f_{\mathcal{R}^+}$ to $\mathcal{ELHI}f$

Lemma 4.1.2. *Let $\mathcal{T}$ be an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox in normal form, and $\mathcal{T}'$ the general TBox obtained after the application of the rule from Table 4.1 has terminated, and A_0, B_0 two concept names occurring in $\mathcal{T}$. Then, $A_0 \sqsubseteq_{\mathcal{T}} B_0$ if and only if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$.*

Proof.

$\Longleftarrow$) We prove by showing the contraposition. Assuming that $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$, then we need to prove that $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$.

Firstly, we prove that every model $\mathcal{I}$ of the TBox $\mathcal{T}$ can be extended to a model $\mathcal{I}'$ for the TBox $\mathcal{T}'$. Given a model $\mathcal{I}$ for the TBox $\mathcal{T}$, we need to check if $\mathcal{I}$ satisfies all new GCIs received by applying the reduction rule in Table 4.1.

> **RR.** Suppose that there are a transitive role r, a GCI $\exists s.A \sqsubseteq B \in \mathcal{T}$ and $r \sqsubseteq s \in \mathcal{T}$ as in the left column of Table 4.1. After **RR** application, we have three new GCIs as in the right column: $\exists r.A \sqsubseteq X$, $X \sqsubseteq B$, $\exists r.X \sqsubseteq X$. We need to prove that the interpretation $\mathcal{I}'$ satisfies all these three GCIs.
>
> Since X is a new concept name that only appears in $\mathcal{T}'$, we define the interpretation of X as follows: $X^{\mathcal{I}'} = (\exists r.A)^{\mathcal{I}'}$. Then we have $\mathcal{I}' \models \exists r.A \sqsubseteq X$. Now we need to prove that $\mathcal{I}' \models X \sqsubseteq B$ and $\mathcal{I}' \models \exists r.X \sqsubseteq X$.
>
> Since $\exists s.A \sqsubseteq B \in \mathcal{T}$ with $r \sqsubseteq s \in \mathcal{T}$, we have $(\exists r.A)^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$. $X^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$ because $X^{\mathcal{I}'} = (\exists r.A)^{\mathcal{I}'}$. Thus $\mathcal{I}' \models X \sqsubseteq B$.
>
> To prove that $\mathcal{I}' \models \exists r.X \sqsubseteq X$, we suppose that $x \in (\exists r.X)^{\mathcal{I}'}$ and then prove that $x \in X^{\mathcal{I}'}$. Since $x \in (\exists r.X)^{\mathcal{I}'}$, there is $y \in X^{\mathcal{I}'}$ such that $(x, y) \in r^{\mathcal{I}'}$. We have $y \in (\exists r.A)^{\mathcal{I}'}$ because $X^{\mathcal{I}'} = (\exists r.A)^{\mathcal{I}'}$. This means that there is $z \in A^{\mathcal{I}'}$ such that $(y, z) \in r^{\mathcal{I}'}$. As $r \circ r \sqsubseteq r$, we have $(x, z) \in r^{\mathcal{I}'}$. Together with $z \in A^{\mathcal{I}'}$, we have $x \in (\exists r.A)^{\mathcal{I}'}$. Thus $x \in X^{\mathcal{I}'}$ because $X^{\mathcal{I}'} = (\exists r.A)^{\mathcal{I}'}$.

We conclude that all the replaced GCIs in $\mathcal{T}'$ are satisfied by $\mathcal{I}$.

Therefore, if $A_0^{\mathcal{I}} \not\subseteq B_0^{\mathcal{I}}$ with $\mathcal{I}$ is a model of $\mathcal{T}$, then $A_0^{\mathcal{I}} \not\subseteq B_0^{\mathcal{I}}$ with $\mathcal{I}$ is a model of $\mathcal{T}'$.

This means that if $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ then $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$, i.e., if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$ then $A_0 \sqsubseteq_{\mathcal{T}} B_0$.

$\Longrightarrow$) We show the contraposition. Assuming that $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$, we need to prove that $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$. $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$ if and only if there is a model $\mathcal{I}'$ of $\mathcal{T}'$ such that there exists $x \in A_0^{\mathcal{I}'}$ but $x \notin B_0^{\mathcal{I}'}$. In order to prove that $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ we build a model $\mathcal{I}$ of $\mathcal{T}$ based on $\mathcal{I}'$ such that there exists $x \in A_0^{\mathcal{I}'}$ but $x \notin B_0^{\mathcal{I}'}$.
We inductively define the interpretation $\mathcal{I}$ as a sequence $\mathcal{I}_0, \ldots, \mathcal{I}_{n_0}$,:
Initially, $\mathcal{I}_0 := \mathcal{I}'$.
After that, for all $n \in \{0, \ldots, n_0 - 1\}$, $\mathcal{I}_{n+1}$ is defined based on $\mathcal{I}_n$

$$\Delta^{\mathcal{I}_{n+1}} := \Delta^{\mathcal{I}_n},$$

$$A^{\mathcal{I}_{n+1}} := A^{\mathcal{I}_n}$$

for all $A \in N_{con}$
The interpretation $\mathcal{I}_{n+1}$ of a role s is defined differently for two cases:

- Case 1: There is no transitive role r such that $r \sqsubseteq_{\mathcal{T}} s$: $s^{\mathcal{I}_{n+1}} := s^{\mathcal{I}_n}$.
- Case 2: There is a transitive role r such that $r \sqsubseteq_{\mathcal{T}} s$, then $s^{\mathcal{I}_{n+1}}$ is repeatedly extended from $s^{\mathcal{I}_n}$:

$$s^{\mathcal{I}_{n+1}} := s^{\mathcal{I}_n} \cup \{(u,w) \mid (u,v),(v,w) \in r^{\mathcal{I}_n}, r \sqsubseteq_{\mathcal{T}} s\}$$

This process of defining $\mathcal{I}$ can be infinite because this DL does not have the finite tree model property. Since $A_0^{\mathcal{I}} = A_0^{\mathcal{I}'}$ and $B_0^{\mathcal{I}} = B_0^{\mathcal{I}'}$, it is readily checked that $x \in A_0^{\mathcal{I}}$ and $x \notin B_0^{\mathcal{I}}$. Before proving that interpretation $\mathcal{I}$ is a model of the general TBox $\mathcal{T}$ we prove a claim:

Claim 4.1.3. If there exist $r \circ r \sqsubseteq r \in \mathcal{T}$, $\exists s.A \sqsubseteq B \in \mathcal{T}$ and $r \sqsubseteq_{\mathcal{T}} s$, then new model $\mathcal{I}$ satisfies axioms: $\exists r.A \sqsubseteq X$, $X \sqsubseteq B$ amd $\exists r.X \sqsubseteq X$ in $\mathcal{T}'$.

Proof.
Since we do not change the interpretations of A, B and X, $X \sqsubseteq B$ holds in $\mathcal{I}$. For $\exists r.A \sqsubseteq X$ and $\exists r.X \sqsubseteq X$, we prove by induction.

- *Base case.* We suppose that $(x,y),(y,z) \in r^{\mathcal{I}'}$, i.e., $(x,y),(y,z) \in r^{\mathcal{I}_0}$ then $(x,z) \in s^{\mathcal{I}_1}$.

Since $r \sqsubseteq_{\mathcal{T}} s$, $\exists s.A \sqsubseteq B \in \mathcal{T}$ and $r \circ r \sqsubseteq r \in \mathcal{T}$, by **RR**, there are $\exists r.A \sqsubseteq X$, $X \sqsubseteq B$, $\exists r.X \sqsubseteq X$ in $\mathcal{T}'$.

$\exists r.A \sqsubseteq X$: Suppose that $z \in A^{\mathcal{I}_1}$, we need to prove that $x \in X^{\mathcal{I}_1}$. Since $A^{\mathcal{I}_1} = A^{\mathcal{I}'}$, we have $z \in A^{\mathcal{I}'}$. Besides, $(y,z) \in r^{\mathcal{I}'}$ and $\exists r.A \sqsubseteq X \in \mathcal{T}'$, thus $y \in X^{\mathcal{I}'}$. Therefore $x \in (\exists r.X)^{\mathcal{I}'}$. Thus by $\exists r.X \sqsubseteq X \in \mathcal{I}'$, we have $x \in X^{\mathcal{I}'}$, i.e. $x \in X^{\mathcal{I}_1}$.

$\exists r.X \sqsubseteq X$: Suppose that $z \in X^{\mathcal{I}_1}$, we need to prove that $x \in X^{\mathcal{I}_1}$. Since $X^{\mathcal{I}_1} = X^{\mathcal{I}'}$, we have $z \in X^{\mathcal{I}'}$. Besides, $(y,z) \in r^{\mathcal{I}'}$ and $\exists r.X \sqsubseteq X \in \mathcal{T}'$, thus $y \in X^{\mathcal{I}'}$. Therefore $x \in (\exists r.X)^{\mathcal{I}'}$. Thus by $\exists r.X \sqsubseteq X \in \mathcal{T}'$, we have $x \in X^{\mathcal{I}'}$, i.e. $x \in X^{\mathcal{I}_1}$.

- *Induction case.* We suppose that $(x,y),(y,z) \in r^{\mathcal{I}_n}$ then $(x,z) \in r^{\mathcal{I}_1}$. Since $\exists r.A \sqsubseteq B \in \mathcal{T}$, by the induction hypothesis, $\mathcal{I}_n \models \exists r.A \sqsubseteq X, X \sqsubseteq B, \exists r.X \sqsubseteq X$.

$\exists r.A \sqsubseteq X$: Suppose that $z \in A^{\mathcal{I}_{n+1}}$, we need to prove that $x \in X^{\mathcal{I}_{n+1}}$. Since $A^{\mathcal{I}_{n+1}} = A^{\mathcal{I}_n}$, we have $z \in A^{\mathcal{I}_n}$. Besides, $(y,z) \in r^{\mathcal{I}_n}$ and $\mathcal{I}_n \models \exists r.A \sqsubseteq X$, thus $y \in X^{\mathcal{I}_n}$. Therefore $x \in (\exists r.X)^{\mathcal{I}_n}$. Thus by $\mathcal{I}_n \models \exists r.X \sqsubseteq X$, we have $x \in X^{\mathcal{I}'}$, i.e. $x \in X^{\mathcal{I}_1}$.

$\exists r.X \sqsubseteq X$: Suppose that $z \in X^{\mathcal{I}_{n+1}}$, we need to prove that $x \in X^{\mathcal{I}_{n+1}}$. Since $X^{\mathcal{I}_{n+1}} = X^{\mathcal{I}_n}$, we have $z \in X^{\mathcal{I}_n}$. Besides, $(y,z) \in r^{\mathcal{I}_n}$ and $\mathcal{I}_n \models \exists r.X \sqsubseteq X$, thus $y \in X^{\mathcal{I}_n}$. Therefore $x \in (\exists r.X)^{\mathcal{I}_n}$. Thus by $\mathcal{I}_n \models \exists r.X \sqsubseteq X$, we have $x \in X^{\mathcal{I}'}$, i.e. $x \in X^{\mathcal{I}_1}$.

Now we prove that the interpretation $\mathcal{I}$ is a model of the general TBox $\mathcal{T}$. We check all the cases of normalized GCIs that appear in $\mathcal{T}$:

- $A \sqsubseteq B$. Given $A^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$, since $A^{\mathcal{I}} = A^{\mathcal{I}'}$ and $B^{\mathcal{I}} = B^{\mathcal{I}'}$, we have $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$.

- $A_1 \sqcap A_2 \sqsubseteq B$. Given $A_1^{\mathcal{I}'} \cap A_2^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$, since $A_1^{\mathcal{I}} = A_2^{\mathcal{I}'}$, $A_2^{\mathcal{I}} = A_2^{\mathcal{I}'}$ and $B^{\mathcal{I}} = B^{\mathcal{I}'}$, we have $A_1^{\mathcal{I}} \cap A_2^{\mathcal{I}} \subseteq B^{\mathcal{I}}$.

- $\exists s.A \sqsubseteq B$. According to the definition of $s^{\mathcal{I}}$, the interpretation of s is extended step by step. We prove that$(\exists s.A)^{\mathcal{I}_n} \subseteq B^{\mathcal{I}_n}$ for all $\mathcal{I}_n$. Given $r \sqsubseteq_{\mathcal{T}} s$, $\exists s.A \sqsubseteq B, \exists s.C \sqsubseteq D \in \mathcal{T}$ and $r \circ r \sqsubseteq r \in \mathcal{T}$. Suppose that at step n, a new element (u, w) is added to $s^{\mathcal{I}_n}$ such that $w \in A^{\mathcal{I}_n}$. We need to prove that $u \in B^{\mathcal{I}_n}$. Since $r \sqsubseteq_{\mathcal{T}} r$, (u, w) is also added to $r^{\mathcal{I}_n}$. In addition, the conditions of Claim 4.1.3 are satisfied, thus, $\mathcal{I}_n$ satisfies $\exists r.A \sqsubseteq X$, $X \sqsubseteq B$, $\exists r.X \sqsubseteq X$. It means that $\mathcal{I}_n$ satisfies $\exists r.A \sqsubseteq B$. Therefore $u \in B^{\mathcal{I}_n}$. Overall, this axiom holds.

- $A \sqsubseteq \exists s.B$. We need to prove that $A^{\mathcal{I}} \subseteq (\exists s.B)^{\mathcal{I}}$. As we only extend the interpretations of roles,i.e., $s^{\mathcal{I}} \supseteq s^{\mathcal{I}'}$. Besides, $A^{\mathcal{I}} = A^{\mathcal{I}'}$, $B^{\mathcal{I}} = B^{\mathcal{I}'}$. Since $A \sqsubseteq \exists s.B \in \mathcal{T}'$, we have $A^{\mathcal{I}'} \subseteq (\exists s.B)^{\mathcal{I}'}$. Therefore $A^{\mathcal{I}} \subseteq (\exists s.B)^{\mathcal{I}}$. This means that this axiom holds.

- $r \sqsubseteq s$. By the definition of the interpretation $\mathcal{I}$, the axiom holds.

- $r \circ r \sqsubseteq r$. This axiom is satisfied by the way we construct $r^{\mathcal{I}}$.

Lemma 4.1.4. *Let $\mathcal{T}$ be an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox in normal form, and $\mathcal{T}'$ the general TBox obtained after the application of the rule from Table 4.1 has terminated. Then the size of $\mathcal{T}'$ is cubic in the size of $\mathcal{T}$.*

4.2 Reduction from $\mathcal{ELHI}f$ to $\mathcal{ELHI}$

In order to eliminate all the axioms $\top \sqsubseteq (\leq 1r)$ in the input general TBox $\mathcal{T}$ we do as follows. If there is an axiom in $\mathcal{T}$ satisfying the left column, then for each concept name $B \in N_{con}$, a new GCI as the one in the right column of Table 4.2 is added to $\mathcal{T}$. When this process terminates, all GCIs of the

Rule	Original axioms	New axioms
RF	$\top \sqsubseteq (\leq 1r)$	$\exists r^{-}.(\exists r.B) \sqsubseteq B$

Table 4.2: Reduction rule from $\mathcal{ELHI}f$ to $\mathcal{ELHI}$

form $\top \sqsubseteq (\leq 1r)$ are eliminated and a new TBox $\mathcal{T}'$ without functional axioms is obtained.

Lemma 4.2.1. *Let $\mathcal{T}$ be an $\mathcal{ELHI}f$ general TBox in normal form, and $\mathcal{T}'$ the general TBox obtained after the application of the rule from Table 4.2 has terminated, and A_0, B_0 two concept names occurring in $\mathcal{T}$. Then, $A_0 \sqsubseteq_{\mathcal{T}} B_0$ if and only if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$.*

Proof.

$\Longleftarrow$) We prove by showing the contraposition: if $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ then $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$.

Firstly, we prove that every model $\mathcal{I}$ of the TBox $\mathcal{T}$ is also a model for the TBox $\mathcal{T}'$. Given a model $\mathcal{I}$ for the TBox $\mathcal{T}$, we need to check if $\mathcal{I}$ satisfies all new GCIs received by applying the reduction rule in Table 4.2.

> **RF.** Suppose that there is a new GCI $\exists r^-.(\exists r.B) \sqsubseteq B$ in $\mathcal{T}'$ and we prove that $\mathcal{I}$ satisfies this new GCI. Assume that we have $z \in (\exists r^-.(\exists r.B))^{\mathcal{I}}$, then we need to prove that $z \in B^{\mathcal{I}}$. Since $z \in (\exists r^-.(\exists r.B))^{\mathcal{I}}$, there is $x \in (\exists r.B)^{\mathcal{I}}$ such that $(x, z) \in r^{\mathcal{I}}$. As $x \in (\exists r.B)^{\mathcal{I}}$, there is $y \in B^{\mathcal{I}}$ such that $(x, y) \in r^{\mathcal{I}}$.
>
> Since both (x, y) and (x, z) are in $r^{\mathcal{I}}$ but $\top \sqsubseteq (\leq 1r)$, we have $z = y$. Thus $z \in B^{\mathcal{I}}$, i.e., I satisfies the new GCI.

We conclude that all the new GCIs in $\mathcal{T}'$ are satisfied by $\mathcal{I}$.

Therefore, if $A_0^{\mathcal{I}} \not\subseteq B_0^{\mathcal{I}}$ with $\mathcal{I}$ is a model of $\mathcal{T}$, then $A_0^{\mathcal{I}} \not\subseteq B_0^{\mathcal{I}}$ with $\mathcal{I}$ is a model of $\mathcal{T}'$.

This means that if $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ then $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$, i.e., if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$ then $A_0 \sqsubseteq_{\mathcal{T}} B_0$ for all concept names A_0 and B_0 in the TBox $\mathcal{T}$.

$\Longrightarrow$) We prove by showing the contraposition: if $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$ then $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$. $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$ if and only if there is a model $\mathcal{I}'$ of $\mathcal{T}'$ such that there exists $x \in A_0^{\mathcal{I}'}$ but $x \notin B_0^{\mathcal{I}'}$. In order to prove that $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ we build a model $\mathcal{I}$ of $\mathcal{T}$ based on $\mathcal{I}'$ such that $x \in A_0^{\mathcal{I}}$ but $x \notin B_0^{\mathcal{I}}$. Based on Theorem 5.6 in [8], it can be assumed that I' is a tree model with multi role names on edges. We first define an interpretation $\mathcal{I}$ of $\mathcal{T}$ such that the condition $x \in A_0^{\mathcal{I}}$ but $x \notin B_0^{\mathcal{I}}$ holds, then prove that $\mathcal{I}$ is a model of $\mathcal{T}$. We inductively define the interpretation $\mathcal{I}$ as a sequence $\mathcal{I}_0, \ldots, \mathcal{I}_{n_0}$,:
Initially, $\mathcal{I}_0 := \mathcal{I}'$.
After that, for all $n \in \{0, \ldots, n_0 - 1\}$, $\mathcal{I}_{n+1}$ is defined based on $\mathcal{I}_n$

$$\Delta^{\mathcal{I}_{n+1}} := \Delta^{\mathcal{I}_n},$$

$$A^{\mathcal{I}_{n+1}} := A^{\mathcal{I}_n}$$

for all $A \in N_{con}$
The interpretation $\mathcal{I}_{n+1}$ of a role s is defined as follows.

> If there are $u, v, w \in \Delta^{\mathcal{I}_n}$ such that $(u, v), (u, w) \in r^{\mathcal{I}_n}$, $v \neq w$ and $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$, then

for all $s \in N_{role}$ such that $(u, w) \in s^{\mathcal{I}_n}$ set:

$$s^{\mathcal{I}_{n+1}} := s^{\mathcal{I}_n} \backslash \{(u, w)\} \cup \{(u, v)\}$$

Else, $s^{\mathcal{I}_{n+1}} := s^{\mathcal{I}_n}$ for all $s \in N_{role}$.

Before defining the interpretation $\mathcal{I}$ we prove a claim.

Claim 4.2.2. Given $(u, v), (u, w) \in r^{\mathcal{I}'}$ such that $v \neq w$ and $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$, then $v \in B^{\mathcal{I}'}$ iff $w \in B^{\mathcal{I}'}$ for all $B \in N_{\mathcal{T}}$.

Proof. We prove only one direction, the other one is totally similar. Suppose that $w \in B^{\mathcal{I}'}$, we need to prove $v \in B^{\mathcal{I}'}$. By **RF** application, there is a GCI $\exists r^{-}.(\exists r.B) \sqsubseteq B$ in $\mathcal{T}'$. Since $(v, u) \in (r^{-})^{\mathcal{I}'}$, $(u, w) \in r^{\mathcal{I}'}$ with $w \in B^{\mathcal{I}'}$, we have $v \in B^{\mathcal{I}'}$.

Now we check all the cases of normalized GCIs that appear in $\mathcal{T}$:

- $A \sqsubseteq B$. We have $A \sqsubseteq B \in \mathcal{T}'$, i.e., $A^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$. Since $A^{\mathcal{I}} = A^{\mathcal{I}'}$ and $B^{\mathcal{I}} = B^{\mathcal{I}'}$, $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$.
- $A_1 \sqcap A_2 \sqsubseteq B$. We have $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}'$, i.e., $A_1^{\mathcal{I}'} \cap A_2^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$. Since $A_1^{\mathcal{I}} = A_2^{\mathcal{I}'}$, $A_2^{\mathcal{I}} = A_2^{\mathcal{I}'}$ and $B^{\mathcal{I}} = B^{\mathcal{I}'}$, $A_1^{\mathcal{I}} \cap A_2^{\mathcal{I}} \subseteq B^{\mathcal{I}}$.
- $\exists r.A \sqsubseteq B$. By the construction of $\mathcal{I}$, a new pair (u, v) can be added to $r^{\mathcal{I}_{n+1}}$. We need to show that this addition does not falsify the axiom $\exists r.A \sqsubseteq B$.

  Suppose that $v \in A^{\mathcal{I}_{n+1}}$, we need to prove that $u \in B^{\mathcal{I}_{n+1}}$.

  By Claim 4.2.2, we have $w \in A^{\mathcal{I}_n}$. By the induction hypothesis, $\mathcal{I}_n \models \exists r.A \sqsubseteq B$, we have $u \in B^{\mathcal{I}_n}$. Thus $u \in B^{\mathcal{I}_{n+1}}$.
- $A \sqsubseteq \exists s.B$. Suppose that $u \in A^{\mathcal{I}_{n+1}}$, we need to prove that $u \in (\exists s.B)^{\mathcal{I}_{n+1}}$.

  We have $\mathcal{I}_n \models A \sqsubseteq \exists s.B$, thus there is $w \in B^{\mathcal{I}_n}$ such that $(u, w) \in s^{\mathcal{I}_n}$.

  There are two possibilities according to the definition of $\mathcal{I}_{n+1}$.

  - If $(u, w) \in r^{\mathcal{I}_{n+1}}$, then $w \in B^{\mathcal{I}_{n+1}}$. Thus $u \in (\exists s.B)^{\mathcal{I}_{n+1}}$, i.e., the axiom holds.
  - Otherwise, $(u, w) \notin s^{\mathcal{I}_{n+1}}$. Since $u \in A^{\mathcal{I}_{n+1}}$, by the definition of $\mathcal{I}_{n+1}$, there is $v \in \Delta^{\mathcal{I}_{n+1}}$ such that $(u, v) \in s^{\mathcal{I}_{n+1}} \backslash s^{\mathcal{I}_n}$. We will prove that $v \in B^{\mathcal{I}_{n+1}}$.

    From the condition of the process extending $\mathcal{I}_n$ to $\mathcal{I}_{n+1}$, there is $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$ such that $(u, v), (u, w) \in r^{\mathcal{I}_n}$.

    From the definition of $r^{\mathcal{I}}$ we see that if $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$ then $r^{\mathcal{I}_n} \subseteq r^{\mathcal{I}'}$ for all n. Thus $(u, v), (u, w) \in r^{\mathcal{I}'}$.

Since $w \in B^{\mathcal{I}_n}$, $w \in B^{\mathcal{I}'}$. By Claim 4.2.2 we have $v \in B^{\mathcal{I}'}$. Together with $v \in \Delta^{\mathcal{I}_{n+1}}$, we conclude that $v \in B^{\mathcal{I}_{n+1}}$.

- $\top \sqsubseteq (\leq 1r)$. Due to the construction of $\mathcal{I}$, all functional roles are satisfied.
- $s_1 \sqsubseteq s_2$. From the definition of $r^{\mathcal{I}}$ and $s^{\mathcal{I}}$ we need to prove for two cases:

  Case 1: A new pair (u, w) with is added to $s_1^{\mathcal{I}_{n+1}}$, then we need to prove that $(u, w) \in s_2^{\mathcal{I}_{n+1}}$. From the definition of $r^{\mathcal{I}_{n+1}}$, there is v such that $(u, v) \in s_1^{\mathcal{I}_n}$ and $(u, v), (u, w) \in r^{\mathcal{I}_n}$ such that $u, v, w \in \Delta^{\mathcal{I}_n}$, $v \neq w$ and $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$. Since $\mathcal{I}_n \models s_1 \sqsubseteq s_2$, $(u, v) \in s_2^{\mathcal{I}_n}$. Therefore by the definition of $s_2^{\mathcal{I}}$ we have $(u, w) \in s^{\mathcal{I}}$.

  Case 2: An old pair $(u, v) \in s_2^{\mathcal{I}_n}$ is removed from $s_2^{\mathcal{I}_n}$. Then by definition of $s_2^{\mathcal{I}_{n+1}}$, $(u, v) \notin s_1^{\mathcal{I}_{n+1}}$.

Lemma 4.2.3. *Let $\mathcal{T}$ be an $\mathcal{ELHI}f$ general TBox in normal form, and $\mathcal{T}'$ the general TBox obtained after the application of the rule from Table 4.2 has terminated. Then the size of $\mathcal{T}'$ is quadratic in the size of $\mathcal{T}$.*

There is a special case of an $\mathcal{ELHI}f$ general TBox without role hierarchies, named $\mathcal{ELI}f$-TBox, that can be reduced to an $\mathcal{ELI}$-TBox whose size is linear in the size of the original one. This DL is considered in the next section.

4.3 Reduction from $\mathcal{ELI}f$ to $\mathcal{ELI}$

In order to eliminate all the axioms $\top \sqsubseteq (\leq 1r)$ in the input TBox $\mathcal{T}$ we do as follows. If there are axioms in $\mathcal{T}$ satisfying the left column, then we *add* new GCIs shown in the right column of Table 4.3. When this process

Rule	Original axioms	New axioms
RF*	$A \sqsubseteq \exists r.B$	$A \sqsubseteq \exists r.\top$
	$\top \sqsubseteq (\leq 1r)$	$\exists r^-.A \sqsubseteq B$

Table 4.3: Reduction rule from $\mathcal{ELI}f$ to $\mathcal{ELI}$

terminates, we eliminate all axioms of the form $\top \sqsubseteq (\leq 1r)$ and get a new TBox $\mathcal{T}'$ without functionality axioms.

Lemma 4.3.1. *Let $\mathcal{T}$ be an $\mathcal{ELI}f$ general TBox in normal form, and $\mathcal{T}'$ the general TBox obtained after the application of the rule from Table 4.3 has terminated, and A_0, B_0 two concept names occurring in $\mathcal{T}$. Then, $A_0 \sqsubseteq_{\mathcal{T}} B_0$ if and only if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$.*

Proof.

$\Longleftarrow$) We show the contraposition, i.e., if $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ then $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$.

Firstly, we prove that every model $\mathcal{I}$ of the TBox $\mathcal{T}$ is also a model for the TBox $\mathcal{T}'$. Given a model $\mathcal{I}$ for the TBox $\mathcal{T}$, we need to check if $\mathcal{I}$ satisfies all new GCIs obtained by applying the reduction rule in Table 4.2.

> **RF*.** Suppose that there is a new GCI $A \sqsubseteq \exists r.\top$ in $\mathcal{T}'$, we have $A \sqsubseteq \exists r.B$, $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$. If there exists $x \in A^{\mathcal{I}}$, then there are $y \in B^{\mathcal{I}}$ such that $(x, y) \in r^{\mathcal{I}}$. This means that $\mathcal{I}$ satisfies $A \sqsubseteq \exists r.\top$. We now need to prove that $\exists r^{-}.A \sqsubseteq B$.
>
> We prove by showing the contradiction. Assume that there is $(u, v) \in r^{\mathcal{I}}$ such that $u \in A^{\mathcal{I}}$ but $v \notin B^{\mathcal{I}}$. As $A \sqsubseteq \exists r.B \in \mathcal{T}$, there is $v' \in B^{\mathcal{I}}$ such that $(u, v') \in r^{\mathcal{I}}$. Since $v \notin B^{\mathcal{I}}$ but $v' \in B^{\mathcal{I}}$, we have $v \neq v'$.
>
> Since both (u, v) and (u, v') with $v \neq v'$ are in $r^{\mathcal{I}}$ but $\top \sqsubseteq (\leq 1r)$, we have the contradiction.
>
> Thus the assumption fails, i.e., the condition holds.

We conclude that all the new GCIs in $\mathcal{T}'$ are satisfied by $\mathcal{I}$.

Therefore, if $A_0^{\mathcal{I}} \not\subseteq B_0^{\mathcal{I}}$ with $\mathcal{I}$ is a model of $\mathcal{T}$, then $A_0^{\mathcal{I}} \not\subseteq B_0^{\mathcal{I}}$ with $\mathcal{I}$ is a model of $\mathcal{T}'$.

This means that if $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ then $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$, i.e., if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$ then $A_0 \sqsubseteq_{\mathcal{T}} B_0$ for all concept names A_0 and B_0 in the TBox $\mathcal{T}$.

$\Longrightarrow$) We prove by showing the contraposition: if $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$ then $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$. $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$ if and only if there is a model $\mathcal{I}'$ of $\mathcal{T}'$ such that there exists $x \in A_0^{\mathcal{I}'}$ but $x \notin B_0^{\mathcal{I}'}$. In order to prove that $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ we build a model $\mathcal{I}$ of $\mathcal{T}$ based on $\mathcal{I}'$ such that $x \in A_0^{\mathcal{I}}$ but $x \notin B_0^{\mathcal{I}}$. By Theorem 5.6 in [8], we can assume that I' is a tree model. First we define an interpretation $\mathcal{I}$ of $\mathcal{T}$ such that the condition $x \in A_0^{\mathcal{I}}$ but $x \notin B_0^{\mathcal{I}}$ holds, and then we prove that $\mathcal{I}$ is a model of $\mathcal{T}$.

We inductively define the interpretation $\mathcal{I}$ as a sequence $\mathcal{I}_0, \ldots, \mathcal{I}_{n_0}$:

Initially, $\mathcal{I}_0 := \mathcal{I}'$.

After that, for all $n \in \{0, \ldots, n_0 - 1\}$, $\mathcal{I}_{n+1}$ is defined based on $\mathcal{I}_n$

$$\Delta^{\mathcal{I}_{n+1}} := \Delta^{\mathcal{I}_n},$$

$$A^{\mathcal{I}_{n+1}} := A^{\mathcal{I}_n}$$

for all $A \in N_{con}$

The interpretation $\mathcal{I}_{n+1}$ of a role r is defined as follows.

> If there are $(u, v), (u, w) \in r^{\mathcal{I}_n}$ such that $v \neq w$ and $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$ then: $r^{\mathcal{I}_{n+1}} := r^{\mathcal{I}_n} \backslash \{(u, w)\}$

Else, $r^{\mathcal{I}_{n+1}} := r^{\mathcal{I}_n}$.

Now we check if $\mathcal{I}$ is a model of $\mathcal{T}$ by checking all the axioms that appear in $\mathcal{T}$.

- $A \sqsubseteq B$. We have $A \sqsubseteq B \in \mathcal{T}'$, i.e.,n $A^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$. Since $A^{\mathcal{I}} = A^{\mathcal{I}'}$ and $B^{\mathcal{I}} = B^{\mathcal{I}'}$, $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$.
- $A_1 \sqcap A_2 \sqsubseteq B$. We have $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}'$, i.e., $A_1^{\mathcal{I}'} \cap A_2^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$. Since $A_1^{\mathcal{I}} = A_2^{\mathcal{I}'}$, $A_2^{\mathcal{I}} = A_2^{\mathcal{I}'}$ and $B^{\mathcal{I}} = B^{\mathcal{I}'}$, $A_1^{\mathcal{I}} \cap A_2^{\mathcal{I}} \subseteq B^{\mathcal{I}}$.
- $\exists r.A \sqsubseteq B$. Since $\exists r.A \sqsubseteq B \in \mathcal{T}'$, we have $(\exists r.A)^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$. As $r^{\mathcal{I}} \subseteq r^{\mathcal{I}'}$ and $A^{\mathcal{I}} = A^{\mathcal{I}'}$, $B^{\mathcal{I}} = B^{\mathcal{I}'}$, we have $(\exists r.A)^{\mathcal{I}} \subseteq B^{\mathcal{I}}$. Thus this axiom holds.
- $A \sqsubseteq \exists r.B$. We prove that this axiom by induction in the definition of $\mathcal{I}$.

  Base case: The axiom holds because $\mathcal{I}_0 = \mathcal{I}'$ and $A \sqsubseteq \exists r.B$ also appears in $\mathcal{T}'$.

  Induction case: There are two possibilities:

  - If $r^{\mathcal{I}_{n+1}} = r^{\mathcal{I}_n}$ then $u \in (\exists r.B)^{\mathcal{I}_{n+1}}$, i.e., the axiom holds.
  - Otherwise, there are $(u, v), (u, w) \in r^{\mathcal{I}_n}$ and $r^{\mathcal{I}_{n+1}} := r^{\mathcal{I}_n} \backslash \{(u, w)\}$. Because of the tree model property, we only consider the case that u is the root of a role. Now assuming that $u \in A^{\mathcal{I}_{n+1}}$, we prove that the elimination of (u, w) does not falsify the axiom.

    If $w \notin B^{\mathcal{I}_n}$ then the axiom holds.

    If $w \in B^{\mathcal{I}_n}$, we prove that $v \in (\exists r.B)^{\mathcal{I}_{n+1}}$ as follows.

    Since $A^{\mathcal{I}_{n+1}} = A^{\mathcal{I}_n}$, $u \in A^{\mathcal{I}_n}$. Due to $A \sqsubseteq \exists r.B \in \mathcal{T}$, by **RF*** we have $\exists r^-.A \sqsubseteq B \in \mathcal{T}'$. Since $(v, u) \in (r^-)^{\mathcal{I}'}$, we have $v \in B^{\mathcal{I}'}$, i.e., $v \in B^{\mathcal{I}_n}$. Since $r^{\mathcal{I}_{n+1}} := r^{\mathcal{I}_n} \backslash \{(u, w)\}$ and $w \neq v$, $(u, v) \in r^{\mathcal{I}_{n+1}}$. Thus $u \in (\exists r.B)^{\mathcal{I}_{n+1}}$.

  Therefore the axiom holds after the inductively defining $\mathcal{I}$.
- $\top \sqsubseteq (\leq 1r)$. We have eliminated all the duplications during we construct $\mathcal{I}$, therefore, all functional roles are satisfied.

Regarding the completion rule in Table 4.3, the maximum number of new GCIs is two times of the number of GCIs of the form $A \sqsubseteq \exists r.B$ appearing in $\mathcal{T}$. Therefore, we give a lemma about the size of the new TBox after the application of this rule has terminated.

Lemma 4.3.2. *Let $\mathcal{T}$ be an $\mathcal{ELI}f$ general TBox in normal form, and $\mathcal{T}'$ the general TBox obtained after the application of the rule from Table 4.3 has terminated. Then the size of $\mathcal{T}'$ is linear in the size of $\mathcal{T}$.*

4.4 Reduction from $\mathcal{ELHI}$ to $\mathcal{ELI}$

In order to eliminate all the role hierarchies $r \sqsubseteq s$ in the input $\mathcal{ELHI}$ -TBox $\mathcal{T}$, we apply the reduction rule in Table 4.4 as follows. We apply Rule **RH** by adding a new GCI as in the right column of Table 4.4 if the GCI and the role hierarchy in its left column are satisfied. When this process terminates,

Rule	Original GCI and RI	New GCIs
RH	$r \sqsubseteq s$ $\exists s.A \sqsubseteq B$	$\exists r.A \sqsubseteq B$

Table 4.4: Reduction rule from $\mathcal{ELHI}$ to $\mathcal{ELI}$

we eliminate all role inclusions and transitive axioms and get a new TBox $\mathcal{T}'$ without role hierarchies.

Lemma 4.4.1. *Let $\mathcal{T}$ be an $\mathcal{ELHI}$ general TBox in normal form, and $\mathcal{T}'$ the general TBox obtained after the application of the rule from Table 4.4 has terminated, and A_0, B_0 two concept names occurring in $\mathcal{T}$. Then, $A_0 \sqsubseteq_{\mathcal{T}} B_0$ if and only if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$.*

Proof.

$\Longleftarrow$) We show the contraposition. Assuming that $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$, then we need to prove that $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$.

Firstly, we prove that every model $\mathcal{I}$ of the TBox $\mathcal{T}$ is also a model for the TBox $\mathcal{T}'$. Given a model $\mathcal{I}$ for the TBox $\mathcal{T}$, we need to check if $\mathcal{I}$ satisfies all new GCIs received by applying the reduction rule in Table 4.4.

> **RH**. If there is a new GCI $\exists r.A \sqsubseteq B$, then there exist $r \sqsubseteq s$ and $\exists s.A \sqsubseteq B$ in $\mathcal{T}$. Since $\mathcal{I} \models r \sqsubseteq s$, $\mathcal{I} \models \exists r.A \sqsubseteq \exists s.A$. Therefore, $\mathcal{I} \models \exists r.A \sqsubseteq B$.

We conclude that all the new GCIs in $\mathcal{T}'$ are satisfied by $\mathcal{I}$.

Therefore, if $A_0^{\mathcal{I}} \not\subseteq B_0^{\mathcal{I}}$ with $\mathcal{I}$ is a model of $\mathcal{T}$, then $A_0^{\mathcal{I}} \not\subseteq B_0^{\mathcal{I}}$ with $\mathcal{I}$ is a model of $\mathcal{T}'$.

This means that if $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ then $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$, i.e., if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$ then $A_0 \sqsubseteq_{\mathcal{T}} B_0$.

$\Longrightarrow$) We prove by showing the contraposition. Assuming that $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$, we need to prove that $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$.

Since $A_0 \not\sqsubseteq_{\mathcal{T}'} B_0$, there is a model $\mathcal{I}'$ of $\mathcal{T}'$ such that there exists $x \in A_0^{\mathcal{I}'}$

but $x \notin B_0^{\mathcal{I}'}$. Based on $\mathcal{I}'$, we define an interpretation $\mathcal{I}$ as follows:

$$\begin{aligned}\Delta^{\mathcal{I}} &:= \Delta^{\mathcal{I}'},\\ A^{\mathcal{I}} &:= A^{\mathcal{I}'},\\ s^{\mathcal{I}} &:= \textstyle\bigcup_{r \sqsubseteq_{\mathcal{T}} s} r^{\mathcal{I}'}.\end{aligned}$$

for all $A \in N_{con}$ and $s \in N_{role}$.

Since $A_0^{\mathcal{I}} = A_0^{\mathcal{I}'}$ and $B_0^{\mathcal{I}} = B_0^{\mathcal{I}'}$, we have $x \in A_0^{\mathcal{I}}$ and $x \notin B_0^{\mathcal{I}}$. Therefore $\mathcal{I} \not\models A_0 \sqsubseteq B_0$.

Now we prove that interpretation $\mathcal{I}$ is a model of the general TBox $\mathcal{T}$. We check all the cases of normalized GCIs that appear in $\mathcal{T}$:

- $A \sqsubseteq B$. Given $A^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$, we have $A^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ because $A^{\mathcal{I}} = A^{\mathcal{I}'}$ and $B^{\mathcal{I}} = B^{\mathcal{I}'}$.
- $A_1 \sqcap A_2 \sqsubseteq B$. Given $A_1^{\mathcal{I}'} \cap A_2^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$, we have $A_1^{\mathcal{I}} \cap A_2^{\mathcal{I}} \subseteq B^{\mathcal{I}}$ because $A_1^{\mathcal{I}} = A_2^{\mathcal{I}'}$, $A_2^{\mathcal{I}} = A_2^{\mathcal{I}'}$ and $B^{\mathcal{I}} = B^{\mathcal{I}'}$.
- $\exists s.A \sqsubseteq B$. Suppose that $x \in (\exists s.A)^{\mathcal{I}}$, we need to prove that $x \in B^{\mathcal{I}}$. We recall the definition of $s^{\mathcal{I}}$, $s^{\mathcal{I}} := \bigcup_{r \sqsubseteq_{\mathcal{T}} s} r^{\mathcal{I}'}$. Since $\mathcal{T} \models r \sqsubseteq s$ and $\exists s.A \sqsubseteq B \in \mathcal{T}$, by **RH** application, there exists $\exists r.A \sqsubseteq B \in \mathcal{T}'$. Therefore, $(\exists r.A)^{\mathcal{I}'} \subseteq B^{\mathcal{I}'}$ for all $r \sqsubseteq_{\mathcal{T}} s$. Together with the fact that $A^{\mathcal{I}} = A^{\mathcal{I}'}$, we have $x \in (\exists s.A)^{\mathcal{I}'}$. Thus $x \in B^{\mathcal{I}'}$. Therefore, if $x \in \bigcup_{r \sqsubseteq_{\mathcal{T}} s} r^{\mathcal{I}'}$, then $x \in B^{\mathcal{I}'}$, i.e., $x \in B^{\mathcal{I}}$.
- $A \sqsubseteq \exists s.B$. Since $A \sqsubseteq \exists s.B \in \mathcal{T}'$, $A^{\mathcal{I}'} \subseteq (\exists s.B)^{\mathcal{I}'}$. As $A^{\mathcal{I}} = A^{\mathcal{I}'}$, $B^{\mathcal{I}} = B^{\mathcal{I}'}$ and $s^{\mathcal{I}} \supseteq s^{\mathcal{I}'}$, we have $A^{\mathcal{I}} \subseteq (\exists s.B)^{\mathcal{I}}$, i.e., the axiom holds.
- $r \sqsubseteq s$. We recall the definitions $r^{\mathcal{I}} := \bigcup_{r' \sqsubseteq_{\mathcal{T}} r} (r')^{\mathcal{I}'}$ and $s^{\mathcal{I}} := \bigcup_{s' \sqsubseteq_{\mathcal{T}} s} (s')^{\mathcal{I}'}$. Since $r \sqsubseteq s \in \mathcal{T}$, we have $r' \sqsubseteq_{\mathcal{T}} s$ for all r' such that $r' \sqsubseteq_{\mathcal{T}} r$. Thus $r^{\mathcal{I}} \subseteq s^{\mathcal{I}}$, i.e., the role inclusion axiom holds.

Lemma 4.4.2. *Let $\mathcal{T}$ be an $\mathcal{ELHI}$ general TBox in normal form, and $\mathcal{T}'$ the general TBox obtained after the application of the rule from Table 4.4 has terminated. Then the size of $\mathcal{T}'$ is quadratic in the size of $\mathcal{T}'$.*

From above reductions, we conclude this chapter by a lemma about the reduction from an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox to an $\mathcal{ELI}$ general TBox.

Lemma 4.4.3. *Let $\mathcal{T}$ be an $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox in normal form, and $\mathcal{T}'$ the $\mathcal{ELI}$ general TBox obtained after applying reductions presented in this chapter, and A_0, B_0 two concept names occurring in $\mathcal{T}$. Then, the size of $\mathcal{T}'$ is polynomial in the size of $\mathcal{T}$ and $A_0 \sqsubseteq_{\mathcal{T}} B_0$ if and only if $A_0 \sqsubseteq_{\mathcal{T}'} B_0$.*

Chapter 5

Classification Algorithms for Intractable Extensions of $\mathcal{EL}$

Algorithms to decide subsumption for extensions of the DL $\mathcal{EL}$ which include role hierarchies, inverse, functional and transitive roles are now developed. From here, we restrict our attention to subsumption between concept names. If we want to check the subsumption between two concept descriptions C and D w.r.t. an input TBox $\mathcal{T}$, we check the subsumption between two new concept names A and B w.r.t. the extended TBox $\mathcal{T}' = \mathcal{T} \cup \{A \sqsubseteq C, D \sqsubseteq B\}$. Then $C \sqsubseteq_{\mathcal{T}} D$ iff $A \sqsubseteq_{\mathcal{T}'} B$.

Now, let $\mathcal{T}$ be a general TBox in normal form, which is defined by an extension of $\mathcal{EL}$, that needs to be classified. Regarding the input TBox $\mathcal{T}$, N_{role} is the set of role names appearing in $\mathcal{T}$, $N_{\mathcal{T}}$ is the union of the set of concept names appearing in $\mathcal{T}$ and $\{\top\}$. To be able to compute the subsumption relations between concept names appearing in $\mathcal{T}$, we define a completion graph for $\mathcal{T}$.

Definition 5.0.4. (The completion graph)
A $\mathcal{T}$-*completion graph* is a tuple (V, E, S), where:

- The set of nodes V is a subset of $N_{\mathcal{T}} \times 2^{\xi}$ with $\xi = \{\exists r.A \mid r \in N_{role}, A \in N_{\mathcal{T}}\}$.
- The set of edges E is a subset of $V \times N_{role} \times V$.
- The labeling function S is a mapping from V to $2^{N_{\mathcal{T}}}$.

Our algorithm builds a completion graph (V, E, S) from $\mathcal{T}$ based on a set of completion rules. The description graph (V, E, S) is initialized as follows.

- $V := \{(A, \emptyset) \mid A \in N_{\mathcal{T}}\}$,
- $S(A, \emptyset) := \{\top, A\}$ for each $A \in N_{\mathcal{T}}$
- $E := \emptyset$.

Then the sets V, E, $S(v)$ for all $v \in V$ are extended by repeatedly applying a set of completion rules until no more rule applies.

The intuition is that the algorithm satisfies two invariants:

- $C \in S(A, \Phi)$ iff $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} C$
- $((A, \Phi), r, (B, \Psi)) \in E$ iff $(A \sqcap \bigsqcap_{\exists s.X \in \Phi} \exists s.X) \sqsubseteq_{\mathcal{T}} \exists r.(B \sqcap \bigsqcap_{\exists s.X \in \Psi} \exists s.X)$

Let (V, E, S) be the completion graph obtained after the application of the set of completion rules for the normalized general TBox $\mathcal{T}$ has terminated. Subsumption between concept names occurring in $\mathcal{T}$ is based on the following relation between subsumption and the completion graph:

$$A \sqsubseteq_{\mathcal{T}} B \text{ iff } B \in S(u) \text{ with } u = (A, \emptyset) \in V$$

where A, B are two concept names occurring in $\mathcal{T}$.

In the following sections we present sets of completion rules that are used by the algorithms to build the completion graph for different intractable $\mathcal{EL}$-extensions. Each section also proves the correctness of the equivalence between subsumption and the completion graph.

We start with the completion rules for $\mathcal{ELI}$ proposed in [29].

5.1 An algorithm for $\mathcal{ELI}$ general TBoxes

In Table 5.1, A, B, A_1, A_2, B_1 are concept names or top; u, v are nodes in the graph; Ψ is a set of concept descriptions of the form $\exists r.A$, and r is either a role or inverse role. In our algorithm, a rule in this table is applied if and only if that application changes the completion graph. The first four rules can be seen as the modification of rules in [12] for the DL $\mathcal{EL}$. The last rule is the one supporting inverse roles and also the one that makes the algorithm not polynomial any more. Intuitively, **CI5** "branches" an edge (u, r, v) to include a new edge (u, r, v') if this new edge effects on the graph. For example, the set of GCIs is

$$\mathcal{T}_0 = \{A \sqsubseteq \exists r.B, \exists r^-.A \sqsubseteq C, \exists r.C \sqsubseteq X\}$$

In the description graph shown in Figure 5.1, after initialization and an application of **CI3**, there is one edge $((A, \emptyset), r, (B, \emptyset))$. However, since there is $\exists r^-.A \sqsubseteq C$ in our set of GCIs, the edge $((A, \emptyset), r, (B, \emptyset))$ is "branched". The label set of the new node $(B, \{\exists r^-.A\})$ has a new element C which has not appeared in the old node $(B, \emptyset)$. Since C is in the label set of $(B, \{\exists r^-.A\})$ and the GCI $\exists r.C \sqsubseteq X$, by **CI4** application, X is added to the label set of $(A, \emptyset)$. Therefore, $A \sqsubseteq_{\mathcal{T}} X$.
The proof that the algorithm terminates and is sound and complete is presented in [29].

CI1	If $A \in S(v)$, $A \sqsubseteq B \in \mathcal{T}$ then $S(v) := S(v) \cup \{B\}$
CI2	If $A_1, A_2 \in S(v)$, $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$ then $S(v) := S(v) \cup \{B\}$
CI3	If $A \in S(u)$, $v = (B, \emptyset)$, $A \sqsubseteq \exists r.B \in \mathcal{T}$ then $E := E \cup \{(u, r, v)\}$
CI4	If $(u, r, v) \in E$, $B \in S(v)$, $\exists r.B \sqsubseteq A \in \mathcal{T}$ then $S(u) := S(u) \cup \{A\}$
CI5	If $(u, r, v) \in E$, $A_1 \in S(u)$, $v = (B, \Psi)$ and $\exists r^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$, $B_1 \notin S(v)$ then $v' := (B, \Psi \cup \{\exists r^-.A_1\})$; if $v' \notin V$ then $V := V \cup \{v'\}$ and $S(v') := S(v) \cup \{B_1\}$, else $S(v') := S(v') \cup \{B_1\}$; $E := E \cup \{(u, r, v')\}$

Table 5.1: Completion Rules for $\mathcal{ELI}$ general TBoxes

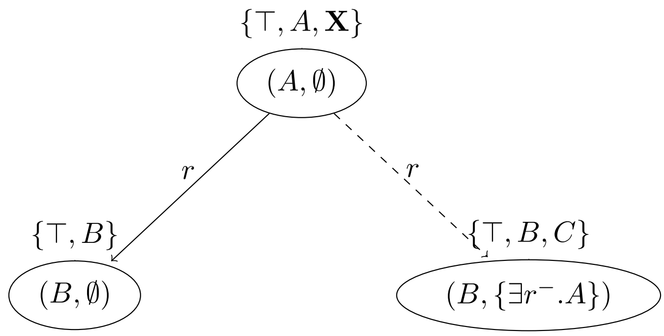

Figure 5.1: An example of reasoning in $\mathcal{ELI}$

We have seen the importance of **CI5**, now we want to extend this algorithm for other intractable extensions of the DL $\mathcal{EL}$.

We have seen in the previous chapter one of the simplest extensions of $\mathcal{ELI}$ is by adding functional roles. There is a linear reduction from an $\mathcal{ELI}f$-TBox to an $\mathcal{ELI}$-TBox presented in the previous chapter. Based on this reduction, the next section presents the algorithm's completion rules for the DL $\mathcal{ELI}f$ as well as a proof of its soundness and completeness.

5.2 An algorithm for $\mathcal{ELI}f$ general TBoxes

Restricted in this section, we define a Boolean function $f : N_{role} \rightarrow \{True, False\}$ such that $f(r) = True$ iff $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$.

The reduction rule from $\mathcal{ELI}f$ to $\mathcal{ELI}$ in Table 4.3 gives us the way to modify the set of completion rules in $\mathcal{ELI}$ for $\mathcal{ELI}f$. In Table 5.2, A, B, A_1, A_2, B_1 are concept names or top; u, v are nodes in the graph; Ψ is a set of concept descriptions having the form $\exists r.A$, and r is either a role or inverse role.

The five rules for $\mathcal{ELI}$ appear in Table 5.2 as **CF1**, **CF2**, the first part of

CF1	If $A \in S(v)$, $A \sqsubseteq B \in \mathcal{T}$ then $S(v) := S(v) \cup \{B\}$
CF2	If $A_1, A_2 \in S(v)$, $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$ then $S(v) := S(v) \cup \{B\}$
CF3	If $A \in S(u)$, $v = (B, \emptyset)$, $w = (\top, \emptyset)$, $A \sqsubseteq \exists r.B \in \mathcal{T}$ then if $f(r)$ then $E := E \cup \{(u, r, w)\}$ else $E := E \cup \{(u, r, v)\}$
CF4	If $(u, r, v) \in E$, $B \in S(v)$, $\exists r.B \sqsubseteq A \in \mathcal{T}$ then $S(u) := S(u) \cup \{A\}$
CF5	If $(u, r, v) \in E$, $B \in S(v)$, $B \sqsubseteq \exists r^-.A$, $f(r^-)$ then $S(u) := S(u) \cup \{A\}$
CF6	If $(u, r, v) \in E$, $A_1 \in S(u)$, $v = (B, \Psi)$ and $\exists r^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$, $B_1 \notin S(v)$ then $v' := (B, \Psi \cup \{\exists r^-.A_1\})$; if $v' \notin V$ then $V := V \cup \{v'\}$; $S(v') := S(v) \cup \{B_1\}$; else $S(v') := S(v') \cup \{B_1\}$; $E := E \cup \{(u, r, v')\}$
CF7	If $(u, r, v) \in E$, $A_1 \in S(u)$, $v = (B, \Psi)$ and $A_1 \sqsubseteq \exists r.B_1 \in \mathcal{T}$, $f(r)$, $B_1 \notin S(v)$ then $v' := (B, \Psi \cup \{\exists r^-.A_1\})$; if $v' \notin V$ then $V := V \cup \{v'\}$; $S(v') := S(v) \cup \{B_1\}$ else $S(v') := S(v') \cup \{B_1\}$; $E := E \cup \{(u, r, v')\}$

Table 5.2: Completion Rules for $\mathcal{ELI}f$ general TBoxes

CF3, **CF4** and **CF6**.

The other rules cater for the consequence of the new axioms from the reduction rule. New GCIs $\exists r^-.A \sqsubseteq B$ and $A \sqsubseteq \exists r.\top$ are added if there are

$A \sqsubseteq \exists r.B$ and $\top \sqsubseteq (\leq 1r)$ in $\mathcal{T}$.

- In **CF3**, if $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$, we extend the graph by an edge from u to $w = (\top, \emptyset)$ instead of $v = (B, \emptyset)$.
- The condition $\exists r.B \sqsubseteq A \in \mathcal{T}$ in **CI4** is replaced by $B \sqsubseteq \exists r^{-}.A \in \mathcal{T}$, $f(r^{-})$ in the new rule **CF5**.
- The condition $\exists r^{-}.A_1 \sqsubseteq B_1 \in \mathcal{T}$ in **CI5** is replaced by $A_1 \sqsubseteq \exists r.B_1 \in \mathcal{T}$,$f(r)$ in the new rule **CF7**.

In the algorithm, a rule in this table is applied if and only if that application changes the completion graph.

For example, the general TBox is

$\mathcal{T}_1 = \{A \sqsubseteq \exists r.B, \exists r^{-}.A \sqsubseteq X, X \sqsubseteq \exists r^{-}.Y, A \sqsubseteq \exists s.C, \top \sqsubseteq (\leq 1r^{-}), \top \sqsubseteq (\leq 1s)\}$

The $\mathcal{T}_1$-completion graph is shown in Figure 5.2

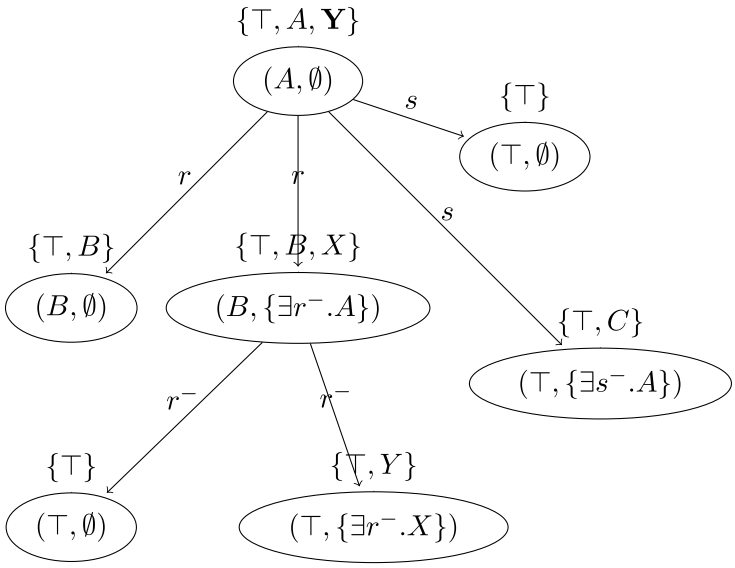

Figure 5.2: An example of reasoning in $\mathcal{ELI}f$

The correctness of the algorithm can be checked by considering the extended TBox as in the reduction rules, then applying the completion rules of $\mathcal{ELI}$.

In the next part we directly prove that the algorithm terminates, is sound and complete. We start with proving termination after exponential time.

Lemma 5.2.1. *For a normalized $\mathcal{ELI}f$ general TBox $\mathcal{T}$, the algorithm runs in exponential time.*

Proof. We need to check the complexity of each rule application and the total number of rule applications.

It is readily checked that rule **CF1** or rule **CF2** application can be performed in polynomial time. The cardinality of V has the upper bound of $|N_{\mathcal{T}}|.2^{|\mathcal{T}|}$ because $V \subseteq N_{\mathcal{T}} \times 2^{\xi}$. Since the cardinality of $N_{\mathcal{T}}$ and ξ are polynomial in the size of $\mathcal{T}$, each application of a rule from **CF3** to **CF7** takes exponential time. Therefore each rule application can be performed in at most exponential time. We now need to count the number of times a rule in Table 5.2 is applied.

Each rule application performed by the algorithm adds at least a new element of $N_{\mathcal{T}}$ to $S(u)$ with $u \in V$, a new edge to E or a new node to V. Since no rule removes any element of the graph, the rules of Table 5.2 can only be applied at most $|N_{\mathcal{T}}|.|V| + |V| + |E|$ times.

The cardinality of E is at most $|N_{role}|.|V|^2$, thus $|E|$ has the upper bound of $|N_{role}|.(|N_{\mathcal{T}}|.2^{|\mathcal{T}|})^2$. Since the cardinality of $N_{\mathcal{T}}$ and N_{role} is linear in the size of $\mathcal{T}$, the total number of rule applications is exponential.

Therefore the algorithm in Section 5.2 runs in exponential time.

Lemma 5.2.2. *(Soundness) Let (V, E, S) be the completion graph obtained after the exhaustive application of the rules from Table 5.2 on the normalized $\mathcal{ELI}f$ general TBox $\mathcal{T}$, and let A_0, B_0 be two concept names occurring in $\mathcal{T}$. Then $A_0 \sqsubseteq_{\mathcal{T}} B_0$ if the following condition holds:*

$$B_0 \in S(u) \text{ with } u = (A_0, \emptyset) \in V$$

Proof. Let $(V_0, E_0, S_0), \ldots, (V_{n_0}, E_{n_0}, S_{n_0})$ be the sequence of description graphs produced by the algorithm. Assume that $B_0 \in S(u)$ with $u = (A_0, \emptyset) \in V$, before proving that $A_0 \sqsubseteq_{\mathcal{T}} B_0$, we prove the following claim.

Claim 5.2.3. Given an $\mathcal{ELI}f$ general TBox $\mathcal{T}$, assume that there are two nodes $u = (A, \Phi)$, $v = (B, \Psi)$, an edge r in the completion graph (V, E, S) and a concept name C.

(a) If $C \in S(A, \Phi)$, then $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} C$; and

(b) if $((A, \Phi), r, (B, \Psi)) \in E$,
then $(A \sqcap \bigsqcap_{\exists s.X \in \Phi} \exists s.X) \sqsubseteq_{\mathcal{T}} \exists r.(B \sqcap \bigsqcap_{\exists s.X \in \Psi} \exists s.X)$

The claim is proved by induction on $n \in \{0, \ldots, n_0 - 1\}$.

For the induction start, $n = 0$ implies $V_n := \{(A, \emptyset) | A \in N_{\mathcal{T}}\}$, $E_n := \emptyset$ and $S_n(v) := \{\top, A\}$ for all $v = (A, \emptyset)$. For (a), $C \in S_0(A, \Phi)$ implies $C = \top$ or $C = A$, thus $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} C$. Since $E_0 = \emptyset$, we have $((A, \Phi), r, (B, \Psi)) \in E$ is always false. Therefore, (b) holds.

For the induction step, we make a case distinction according to the rule used to add a new concept to $S_n(v)$ for $v \in V_n$ or new edge to E_n. Within those seven rules, there are **CF1**, **CF2**, **CF4** and **CF5** that can add new concept names to $S_n(v)$. Thus, it is necessary to prove that condition (a) still holds when applying those four rules:

CF1 Suppose that after applying this rule, $S_{n+1}(v) = S_n(v) \cup B$ with $v = (A^*, \Phi)$. This means that there exist $A \in S_n(v)$ and $A \sqsubseteq B \in \mathcal{T}$. By induction hypothesis, we have $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} A$, thus $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} B$.

CF2 Suppose that after applying this rule, $S_{n+1}(v) = S_n(v) \cup B$ with $v = (A^*, \Phi)$. This means that there exist $A_1, A_2 \in S_n(v)$ and $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$. We have $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} A_1$ and $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} A_2$, thus $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} B$.

CF4 Suppose that after applying this rule, $S_{n+1}(u) = S_n(u) \cup A$ with $u = (A^*, \Omega)$. This means that there exist $(u, r, v) \in E_n$ with $v = (B^*, \Phi)$, $B \in S_n(v)$ and $\exists r.B \sqsubseteq A \in \mathcal{T}$. From invariant (b) and induction hypothesis, we have $(A^* \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.(B^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X)$, and from (a) at step n, we get $(B^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} B$. Thus $(A^* \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.B$. Together with $\exists r.B \sqsubseteq A \in \mathcal{T}$, we get the conclusion $(A^* \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X) \sqsubseteq_{\mathcal{T}} A$, preserving invariant (a) as required.

CF5 Suppose that after applying this rule, $S_{n+1}(u) = S_n(u) \cup A$ with $u = (A^*, \Omega)$. This means that there exist $(u, r, v) \in E_n$ with $v = (B^*, \Phi)$, $B \in S_n(v)$ and $B \sqsubseteq \exists r^-.A \in \mathcal{T}$, $f(r^-)$. Given an interpretation $\mathcal{I}$ of $\mathcal{T}$, suppose that there is an $x \in (A^* \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X)^{\mathcal{I}}$, from condition (b), there is a $y \in (B^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X)^{\mathcal{I}}$ such that $(x, y) \in r^{\mathcal{I}}$. We have $x \in A^{\mathcal{I}}$, because $(y, x) \in (r^-)^{\mathcal{I}}$ and $f(r^-)$. Thus we get the conclusion $(A^* \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X) \sqsubseteq_{\mathcal{T}} A$

In the set of rules in Table 5.2, **CF3** may add a new edge to the completion graph. Thus we need to prove that (b) still holds after applying this rule.

CF3 Suppose that after applying this rule, $E_{n+1} = E_n \cup (u, r, v)$. This means that there exists $u = (A^*, \Phi)$, and $A \in S_n(u)$, $v = (B, \emptyset)$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$. From (a), $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} A$, thus $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.B$. Since $v = (B, \emptyset)$, we see that (b) holds.

Suppose that after applying this rule, $E_{n+1} = E_n \cup (u, r, w)$. This means that there exists $u = (A^*, \Phi)$, and $A \in S_n(u)$, $w = (\top, \emptyset)$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$. Since $A \sqsubseteq \exists r.B \in \mathcal{T}$, $A \sqsubseteq_{\mathcal{T}} \exists r.\top$. From (a), $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} A$, thus $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.\top$. Since $w = (\top, \emptyset)$, we see that (b) holds.

In the set of rules in Table 5.2, **CF6** and **CF7** add new elements to both E_n and S_n. Thus we need to prove that both (a) and (b) still hold after applying this rule.

CF6 We prove that both the conditions hold.

(a) holds.
If $v' \in V_n$, suppose that after applying this rule, $S_{n+1}(v') := S_n(v') \cup \{B_1\}$. Since $v' := (B, \Psi \cup \{\exists r^-.A_1\})$ and $\exists r^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$, (a) holds.
If $v' \notin V_n$, then $S_{n+1}(v') := S_n(v) \cup \{B_1\}$. Since $v' := (B, \Psi') = (B, \Psi \cup \{\exists r^-.A_1\})$, we have $(B \sqcap \bigsqcap_{\exists t.X \in \Psi'} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r^-.A_1$. Thus $(B \sqcap \bigsqcap_{\exists t.X \in \Psi'} \exists t.X) \sqsubseteq_{\mathcal{T}} B_1$, i.e., (a) holds.

(b) holds.
In both cases ($v' \in V_n$) and ($v' \notin V_n$), Rule CR5 computes $E_{n+1} = E_n \cup \{(u, r, v')\}$. Therefore we prove that (b) holds for two cases at the same time.
By induction hypothesis and $(u, r, v) \in E_n$, we have $(A \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.(B \sqcap \bigsqcap_{\exists t.X \in \Psi'} \exists t.X)$. Given an $x \in (A \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X)^{\mathcal{I}}$ with $\mathcal{I}$ as a model of $\mathcal{T}$, there is a $y \in (B \sqcap \bigsqcap_{\exists t.X \in \Psi} \exists t.X)^{\mathcal{I}}$, such that $(x, y) \in r^{\mathcal{I}}$, i.e., $(y, x) \in r^{-\mathcal{I}}$. By (a), we have $x \in A_1^{\mathcal{I}}$. Hence $y \in (\exists r^-.A_1)^{\mathcal{I}}$. Thus $y \in (B \sqcap \exists r^-.A_1 \sqcap \bigsqcap_{\exists t.X \in \Psi} \exists t.X)^{\mathcal{I}} = (B \sqcap \bigsqcap_{\exists t.X \in \Psi'} \exists t.X)^{\mathcal{I}}$. Therefore, $(A \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.(B \sqcap \bigsqcap_{\exists t.X \in \Psi'} \exists t.X)$, i.e., (b) holds.

CF7 First we prove that if $A_1 \sqsubseteq \exists r.B_1 \in \mathcal{T}$, $f(r)$ then $\exists r^-.A_1 \sqsubseteq_{\mathcal{T}} B_1$. Suppose that $y \in (\exists r^-.A_1)^{\mathcal{I}}$, then we need to prove that $y \in B_1^{\mathcal{I}}$. There is an $x \in A_1^{\mathcal{I}}$ such that $(x, y) \in r^{\mathcal{I}}$ because $y \in (\exists r^-.A_1)^{\mathcal{I}}$. Since $A_1 \sqsubseteq \exists r.B_1 \in \mathcal{T}$, there is $y' \in B_1^{\mathcal{I}}$ such that $(x, y') \in r^{\mathcal{I}}$. However $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$, we have $y = y'$, i.e., $y \in B_1^{\mathcal{I}}$.

Then we use the proof of **CF6** to prove that both the conditions hold after **CF7** application.

We have finished the proof of Claim 5.2.3.

Using the Claim 5.2.3, it is now easy to prove that $A_0 \sqsubseteq_{\mathcal{T}} B_0$. Let $B_0 \in S(A_0, \emptyset)$, by point (a) of Claim 5.2.3, we have $(A_0 \sqcap \bigsqcap_{\exists t.X \in \emptyset} \exists t.X) \sqsubseteq_{\mathcal{T}} B_0$, i.e., $A_0 \sqsubseteq_{\mathcal{T}} B_0$.

Lemma 5.2.4. *(Completeness) Let (V, E, S) be the completion graph obtained after the exhaustive application of the rules from Table 5.2 for the normalized $\mathcal{ELIf}$ general TBox $\mathcal{T}$, and let A_0, B_0 be concept names occurring in $\mathcal{T}$. Then $A_0 \sqsubseteq_{\mathcal{T}} B_0$ implies that there exists a node $u = (A_0, \emptyset) \in V$ such that:*

$$B_0 \in S(u)$$

Proof. The lemma is proved by showing the contraposition. Suppose that $B_0 \notin S(u)$ with $u = (A_0, \emptyset)$, we need to prove that this implies $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$. We prove $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ by giving a counter model $\mathcal{I}$ for the general

TBox $\mathcal{T}$, in which there is a witness $x \in \Delta^{\mathcal{I}}$ such that $x \in A_0^{\mathcal{I}}$ but $x \notin B_0^{\mathcal{I}}$. Before actually defining this model, we define a set E_{bad} of bad edges and prove a claim.

Given a completion graph (V, E, S), an edge $(u, r, v) \in E$ belongs to the *bad edge set* E_{bad} if we can find $\exists r^-.A \sqsubseteq B \in \mathcal{T}$ or both $A \sqsubseteq \exists r.B, \top \sqsubseteq (\leq 1r) \in \mathcal{T}$, such that $A \in S(u)$ and $B \notin S(v)$.

In our example in Figure 5.2, we have

$$E_{bad} = \{((A, \emptyset), r, (B, \emptyset)), ((A, \emptyset), s, (\top, \emptyset)), ((B, \exists r^-.A), r^-, (\top, \emptyset))\}$$

Claim 5.2.5. Given the completion graph (V, E, S) and the bad edge set E_{bad}, we have:

If $(u, r, v) \in E_{bad}$ with $v = (A, \Phi)$, then there is
$(u, r, v') \in E \backslash E_{bad}$ with $v' = (A, \Psi)$ and $\Phi \subsetneq \Psi$

The claim is proved as follows. Let (u, r, v) with $v = (A, \Phi)$ be a bad edge. This means that there is $\exists r^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$ with $A_1 \in S(u)$ and $B_1 \notin S(v)$,

Firstly, we prove that $\exists r^-.A_1 \notin \Phi$ by contradiction. Node v is "branched" from a node $v_0 = (A, \emptyset)$, then it is extended by **CF6** and **CF7** applications. If $\exists r^-.A_1 \in \Phi$, during the process of rule applications that v_0 is branched to v, there is a step that a node (A, Φ') with $\Phi' \subseteq \Phi \backslash \{\exists r^-.A_1\}$. After **CF6** or **CF7** application on (A, Φ'), we have $B_1 \in S(v)$, which is contrary to the hypothesis.

Then we prove that when the algorithm terminates, there is a node (A, Ψ) and an edge $(u, r, (A, \Psi))$ such that there is not any $\exists r^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$ or both $A \sqsubseteq \exists r.B, \top \sqsubseteq (\leq 1r) \in \mathcal{T}$ such that $A_1 \in S(u)$ and $B_1 \notin S(A, \Psi)$ and $\exists r^-.A_1 \notin \Psi$. By the definition of E_{bad}, we have $(u, r, (A, \Psi)) \notin E_{bad}$.

That conclusion is proved by checking **CF6** and **CF7** applications.

If $v' = (A, \Phi \cup \{\exists r^-.A_1\}) \in V$, then after **CF6** or **CF7** application, we have $B_1 \in S(v')$ and $(u, r, v') \in E$. Otherwise, if $v' \notin V$, then after **CF6** or **CF7** application, this new node v' with $B_1 \in S(v')$ and a new edge (u, r, v') are added to the graph.

That process is repeated until there is $(A, \Psi) \in V$ with $\Phi \cup \{\exists r^-.A_1\} \subseteq \Psi$ and $(u, r, (A, \Psi)) \in E$ such that there is not any $\exists r^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$ or both $A \sqsubseteq \exists r.B, \top \sqsubseteq (\leq 1r) \in \mathcal{T}$ such that $A_1 \in S(u)$ and $B_1 \notin S(v')$.

Therefore when the algorithm terminates there is a node $v' = (A, \Psi)$ with $\Phi \subsetneq \Psi$ such that (u, r, v') is not a bad edge.

Regarding the functional property, from a node u and a role name r, we pick only one edge in our graph by defining a function.

$$\tau : V \times N_{role} \rightarrow V$$

Given $u \in V$ and $r \in N_{role}$, then $\tau(u,r)$ is an element in the set $\{v \mid (u,r,v) \in E \backslash E_{bad}$.

We inductively define a sub-graph (V^*, E^*, S) of V, E, S as follows.

$V_0^* = \{(A_0, \emptyset)\}$;

If there is $(u,r,v) \in E \backslash E_{bad}$ such that $u \in V_i^*$ then

if $\neg f(r)$ then $V_{i+1}^* := V_i^* \cup \{v\}$; $E_{i+1}^* := E_i^* \cup \{(u,r,v)\}$

else if $f(r)$ and there is $(v', r^-, u) \in E_i^*$ then $V_{i+1}^* := V_i^*$; $E_{i+1}^* := E_i^* \cup \{(u,r,v')\}$

else $V_{i+1}^* := V_i^* \cup \{\tau(u,r)\}$; $E_{i+1}^* := E_i^* \cup \{(u,r,\tau(u,r))\}$

until the defining graph cannot change.

In Figure 5.2, we have

$E^* = \{((A,\emptyset), r, (B, \exists r^-.A)), ((A,\emptyset), s, (\top, \exists s^-.A)), ((B, \exists r^-.A), r^-, (\top, \exists r^-.X))\}$

We now define a model $\mathcal{I}$ based on (V^*, E^*, S) as follows:

$$
\begin{aligned}
\Delta^{\mathcal{I}} &:= V^*; \\
A^{\mathcal{I}} &:= \{u \mid A \in S(u), u \in V^*\}; \\
r^{\mathcal{I}} &:= \{(u,v) \mid (u,r,v) \in E^*\} \cup \{(v,u) \mid (u,r^-,v) \in E^*\}.
\end{aligned}
$$

for all $A \in N_{con}$ and $r \in N_{role}$.

Then we unravel this model with the root at $(A_0, \emptyset)$ and get a possibly infinite tree model $\mathcal{I}$.

In the example, we call $u = (A,\emptyset), v = (B, \exists r^-.A), w = (\top, \exists r^-.A)$. The interpretation $\mathcal{I}$ is:

$\Delta^{\mathcal{I}} = \{u, v, w\}$;

$A^{\mathcal{I}} = \{u\}$; $B^{\mathcal{I}} = \{v\}$; $C^{\mathcal{I}} = \{w\}$; $X^{\mathcal{I}} = \{v\}$; $Y^{\mathcal{I}} = \{u\}$;

$r^{\mathcal{I}} = \{(u,v)\}; s^{\mathcal{I}} = \{(v,w)\}$.

First we show that there is an $x \in \Delta^{\mathcal{I}}$ such that $x \in A_0^{\mathcal{I}} \backslash B_0^{\mathcal{I}}$, then we prove that the interpretation $\mathcal{I}$ is a model of our general TBox $\mathcal{T}$. Considering the node $u = (A_0, \emptyset) \in \Delta^{\mathcal{I}}$ we have $u \notin B_0^{\mathcal{I}}$ since $B_0 \notin S(u)$ by the supposed hypothesis. Furthermore, the algorithm starts with $S(A, \emptyset) := \{\top, A\}$ for each $A \in N_{\mathcal{T}}$. So, $A_0 \in S(u)$, implying that $u \in A_0^{\mathcal{I}}$. Therefore, $u = (A_0, \emptyset) \in (A_0^{\mathcal{I}} \backslash B_0^{\mathcal{I}})$.

Now we need to prove that $\mathcal{I}$ is a model of $\mathcal{T}$. We make a case distinction according to the form of concept inclusions.

- $A \sqsubseteq B$.
  Let $u \in A^{\mathcal{I}}$, by the definition of $A^{\mathcal{I}}$, we have $A \in S(u)$. Due to Rule **CF1**, this implies $B \in S(u)$, thus $u \in B^{\mathcal{I}}$.

- $A_1 \sqcap A_2 \sqsubseteq B$.
  Let $u \in A_1^{\mathcal{I}} \sqcap A_2^{\mathcal{I}}$, by the definition of $A_1^{\mathcal{I}}$, $A_2^{\mathcal{I}}$, we have $A_1, A_2 \in S(u)$. Due to Rule **CF2**, this implies $B \in S(u)$, thus $u \in B^{\mathcal{I}}$.

- $A \sqsubseteq \exists r.B$.
  Let $u \in A^{\mathcal{I}}$, by the definition of $A^{\mathcal{I}}$, we have $A \in S(u)$.

  Due to **CF3** and the definition of $\mathcal{I}$, there exists $(u, r, v) \in E$ with $v = (B, \emptyset)$.

  If r is not a functional role, then there are two cases:

  - *First case:* $(u, r, v) \in E \backslash E_{bad}$: From the definition of $r^{\mathcal{I}}$, we have $(u, v) \in r^{\mathcal{I}}$. Because $B \in S(B, \emptyset)$, we have $v \in B^{\mathcal{I}}$. Together, this yields $u \in (\exists r.B)^{\mathcal{I}}$.
  - *Second case:* $(u, r, v) \in E_{bad}$: From Claim 5.2.5, there is $v' = (B, \Phi) \neq v$ such that $(u, r, v') \in E \backslash E_{bad}$. Therefore $(u, v') \in r^{\mathcal{I}}$. Because $B \in S(B, \Phi)$ for all Φ, we have $v' \in B^{\mathcal{I}}$. Together, this yields $u \in (\exists r.B)^{\mathcal{I}}$.

  If r is a functional role, we need to check two cases

  - If there is $u' \in \Delta^{\mathcal{I}}$ such that $(u', u) \in (r^-)^{\mathcal{I}}$, i.e., $(u', r^-, u) \in E$: by **CF5** application, we have $B \in S(u')$, i.e., $u' \in B^{\mathcal{I}}$. Since $(u, u') \in r^{\mathcal{I}}$, this axiom holds.
  - If there is not any $u' \in \Delta^{\mathcal{I}}$ such that $(u', u) \in (r^-)^{\mathcal{I}}$: then $(u, \tau(u, r)) \in r^{\mathcal{I}}$. Therefore, this axiom holds.

- $\exists r.A \sqsubseteq B$.
  Let $u \in (\exists r.A)^{\mathcal{I}}$, which means that there is a $v \in A^{\mathcal{I}}$ such that $(u, v) \in r^{\mathcal{I}}$. According to the definition of $r^{\mathcal{I}}$ above, there are two possibilities:

  - *First case:* $(u, r, v) \in E \backslash E_{bad}$: Due to Rule **CF4**, we have $B \in S(u)$, which means that $u \in B^{\mathcal{I}}$.
  - *Second case:* $(v, r^-, u) \in E \backslash E_{bad}$: As we know the relation between r and r^-:

$$(u, v) \in r^{\mathcal{I}} \text{ iff } (v, u) \in (r^-)^{\mathcal{I}}$$

We rename $s = r^-$, i.e., $r = s^-$ to make its look more appropriate:
$(v, s, u) \in E \backslash E_{bad}$, $\exists s^-.A \sqsubseteq B \in \mathcal{T}$, $A \in S(v)$,
Together with the fact that $(v, s, u) \notin E_{bad}$, we have $B \in S(u)$, which means that $u \in B^{\mathcal{I}}$.

- $\top \sqsubseteq (\leq 1r)$.
  This axiom holds by the way we construct the interpretation.

Thus the completeness of the algorithm is proved.

5.3 An algorithm for $\mathcal{ELHI}_{\mathcal{R}^+}$ general TBoxes

The set of completion rules in Table 5.3 is used to construct the description graph. The first five rules in Table 5.3 are derived from the completion rules for $\mathcal{EL}^+$ in Fig. 2. in [9]. The difference is that here some information about roles is implicitly represented in hierarchy axioms.

The last two rules are "branching" rules.

- **CR6** is similar to **CI5**, whose condition includes $\exists s^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$ and $B_1 \notin S(v)$.
- **CR7** is provided to support transitive roles.

In Table 5.3, A, B, A_1, A_2 and B_1 are concept names or top; u, v are nodes in the graph; Φ and Ψ are sets of concept descriptions of the form $\exists r.A$, and $r, r_1, r_1^-, r_2, s, s^-$ are either roles or inverse roles.

In our algorithm, a rule in the table is applied if and only if that application changes the completion graph. In the next part we prove that the algorithm terminates, is sound and complete.

We consider an example TBox in Figure 5.3 to demonstrate the algorithm.

$A \sqsubseteq \exists r_1.B$	$r_1 \sqsubseteq r$	$r \circ r \sqsubseteq r$
$B \sqsubseteq \exists r_2.C$	$r_2 \sqsubseteq r$	
$\exists s^-.A \sqsubseteq D$	$r \sqsubseteq s$	
$\exists s.D \sqsubseteq X$		

Figure 5.3: An $\mathcal{ELHI}_{\mathcal{R}^+}$ general TBox

Using the completion rules in Table 5.3, a completion graph is built as in Figure 5.4. After initialization and **CR3, CR5** applications we have the

CR1	If $A \in S(v)$, $A \sqsubseteq B \in \mathcal{T}$ then $S(v) := S(v) \cup \{B\}$
CR2	If $A_1, A_2 \in S(v)$, $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$ then $S(v) := S(v) \cup \{B\}$
CR3	If $A \in S(u)$, $v = (B, \emptyset)$, $A \sqsubseteq \exists r.B \in \mathcal{T}$ then $E := E \cup \{(u, r, v)\}$
CR4	If $(u, r, v) \in E$, $B \in S(v)$, $\exists s.B \sqsubseteq A \in \mathcal{T}$, $r \sqsubseteq_{\mathcal{T}} s$ then $S(u) := S(u) \cup \{A\}$
CR5	If $(u, r_1, v), (v, r_2, w) \in E$, $r_1 \sqsubseteq_{\mathcal{T}} s$, $r_2 \sqsubseteq_{\mathcal{T}} s$, $s \circ s \sqsubseteq s \in \mathcal{T}$ then $E := E \cup \{(u, s, w)\}$
CR6	If $(u, r, v) \in E$, $v = (B, \Psi)$, $\exists s^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$, $r \sqsubseteq_{\mathcal{T}} s$, and $A_1 \in S(u)$ then $v' := (B, \Psi \cup \{\exists r^-.A_1\})$; if $v' \notin V$ then $V := V \cup \{v'\}$ and $S(v') := S(v) \cup \{B_1\}$, else $S(v') := S(v') \cup \{B_1\}$; $E := E \cup \{(u, r, v')\}$
CR7	If $(u, r_2, v) \in E$, $u = (A, \Phi)$, $v = (B, \Psi)$, $r \circ r \sqsubseteq r \in \mathcal{T}$ and $r_1 \sqsubseteq_{\mathcal{T}} r$, $r_2 \sqsubseteq_{\mathcal{T}} r$, $\exists r_1^-.A_1 \in \Phi$, $\exists s^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$, $r \sqsubseteq_{\mathcal{T}} s$ then $v' := (B, \Psi \cup \{\exists r^-.A_1\})$; if $v' \notin V$ then $V := V \cup \{v'\}$ and $S(v') := S(v) \cup \{B_1\}$, else $S(v') := S(v') \cup \{B_1\}$; $E := E \cup \{(u, r_2, v')\}$

Table 5.3: Completion Rules for $\mathcal{ELHI}_{\mathcal{R}^+}$ general TBoxes

sub-graph of Figure 5.4 with only three nodes u, v, w. Then two applications of **CR6** create two new nodes v', w' and two edges $(u, v'), (u, w')$. **CR7** application creates the edge (v', r_2, w'). Lastly, **CR4** adds X to the label set of u. Thus, we have $A \sqsubseteq X$ w.r.t. the general TBox in Figure 5.3 as required.
We start with proving termination after exponential time.

Lemma 5.3.1. *For a normalized $\mathcal{ELHI}_{\mathcal{R}^+}$ general TBox $\mathcal{T}$, the algorithm runs in exponential time.*

Proof. We need to check the complexity of each rule application and the total number of rule applications.

It is readily checked that rule **CR1** or rule **CR2** application can be performed in polynomial time. Since $V \subseteq N_{\mathcal{T}} \times 2^{\xi}$, the cardinality of V has the upper bound of $|N_{\mathcal{T}}|.2^{|\xi|}$. Since the cardinality of $N_{\mathcal{T}}$ is linear in

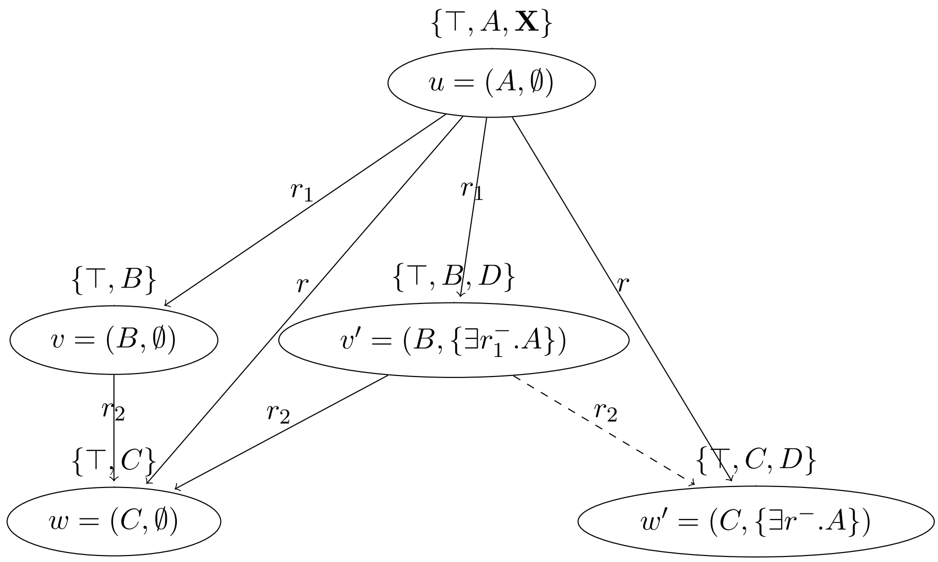

Figure 5.4: An example of reasoning in $\mathcal{ELHI}_{\mathcal{R}^+}$

the size of $\mathcal{T}$, of ξ is quadratic in the size of $\mathcal{T}$, each application of a rule from **CR3** to **CR7** takes exponential time. Therefore each rule application can be performed in at most exponential time. We now need to count the number of times a rule in Table 5.3 is applied.

Each rule application performed by the algorithm adds at least a new element of $N_{\mathcal{T}}$ to $S(u)$ with $u \in V$, a new edge to E or a new node to V of the completion graph. Since no rule removes any element of the graph, the rules of Table 5.3 can only be applied at most $|N_{\mathcal{T}}|.|V| + |V| + |E|$ times.

The cardinality of E is at most $|N_{role}|.|V|^2$, thus $|E|$ has the upper bound of $|N_{role}|.(|N_{\mathcal{T}}|.2^{|\mathcal{T}|})^2$. Due to the fact that the cardinality of $N_{\mathcal{T}}$ and N_{role} is linear in the size of $\mathcal{T}$, the total number of rule applications is exponential.

Therefore the algorithm in Section 5.3 runs in exponential time.

Lemma 5.3.2. *(Soundness) Let (V, E, S) be the completion graph obtained after the application of the rules from Table 5.3 on the normalized general TBox $\mathcal{T}$ has terminated, and let A_0, B_0 be two concept names occurring in $\mathcal{T}$. Then $A_0 \sqsubseteq_{\mathcal{T}} B_0$ if the following condition holds:*

$$B_0 \in S(u) \textit{ with } u = (A_0, \emptyset) \in V$$

Proof. Let $(V_0, E_0, S_0), \ldots, (V_{n_0}, E_{n_0}, S_{n_0})$ be the sequence of description graphs produced by the algorithm. Suppose that $B_0 \in S(u)$ with $u = (A, \emptyset) \in V$. Before proving that $A \sqsubseteq_{\mathcal{T}} B$, we prove the following claim.

Claim 5.3.3. Given the general TBox $\mathcal{T}$, assume that there are two nodes $u = (A, \Phi)$, $v = (B, \Psi)$, an edge r in the completion graph (V, E, S) and a concept name C.

(a) If $C \in S(A, \Phi)$, then $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} C$; and

(b) if $((A, \Phi), r, (B, \Psi)) \in E$,
then $(A \sqcap \bigsqcap_{\exists s.X \in \Phi} \exists s.X) \sqsubseteq_{\mathcal{T}} \exists r.(B \sqcap \bigsqcap_{\exists s.X \in \Psi} \exists s.X)$

The claim is proved by induction on $n \in \{0, \dots, n_0 - 1\}$.

For the induction start, $n = 0$ implies $S_n(v) := \{\top, A\}$ for all $v = (A, \emptyset)$, $V_n := \{(A, \emptyset) | A \in N_{\mathcal{T}}\}$, $E_n := \emptyset$. For (a), $C \in S_0(A, \Phi)$ implies $C = \top$ or $C = A$, thus $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} C$. Since $E_0 = \emptyset$, $((A, \Phi), r, (B, \Psi)) \in E$ is always false. Thus (b) holds.

For the induction step, we make a case distinction according to the rule used to add a new concept to $S_n(v)$ for $v \in V_n$ or new edge to E_n. Within those seven rules, there are **CR1**, **CR2** and **CR4** that can add new concept names to $S_n(v)$. Thus, it is needed to prove that condition (a) still holds when applying those three rules:

CR1 and **CR2** We prove in the same way as the proofs of **CF1, CF2** in Claim 5.2.3.

CR4 Suppose that after applying this rule, $S_{n+1}(u) = S_n(u) \cup A$ with $u = (A^*, \Phi)$. This means that there exist $(u, r, v) \in E_n$ with $v = (B^*, \Phi)$, $B \in S_n(v)$ and $\exists s.B \sqsubseteq A \in \mathcal{T}$, $r \sqsubseteq_{\mathcal{T}} s$. Since $\exists s.B \sqsubseteq A \in \mathcal{T}$, $r \sqsubseteq_{\mathcal{T}} s$, we have $\exists r.B \sqsubseteq_{\mathcal{T}} A$. From condition (b), we have $(A^* \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.(B^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X)$, and from (a) at step n, we get $(B^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} B$. Thus $(A^* \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.B$. Together with $\exists r.B \sqsubseteq_{\mathcal{T}} A$, we get the conclusion $(A^* \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X) \sqsubseteq_{\mathcal{T}} A$

In the set of rules in Table 5.3, **CR3** and **CR5** add elements to the edge set E_n . So we need to prove that (b) still holds after applying one of those three rules.

CR3 Suppose that after applying this rule, $E_{n+1} = E_n \cup (u, r, v)$. This means that there exists $u = (A^*, \Phi)$, and $A \in S_n(u)$, $v = (B, \emptyset)$ and $A \sqsubseteq \exists r.B \in \mathcal{T}$. From (a), $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} A$, thus $(A^* \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r.B$. Since $v = (B, \emptyset)$, we see that (b) holds.

CR5 Suppose that after applying this rule, $E_{n+1} = E_n \cup (u, s, w)$. This means that there exist $(u, r_1, v), (v, r_2, w) \in E_n$ and $s \circ s \sqsubseteq s \in \mathcal{T}$. Assume that $u = (A, \Phi)$, $v = (B, \Psi)$ and $w = (C, \Omega)$, we have $(A \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r_1.(B \sqcap \bigsqcap_{\exists t.X \in \Psi} \exists t.X)$ and $(B \sqcap \bigsqcap_{\exists t.X \in \Psi} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists r_2.(C \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X)$. If there is $x \in (A \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X)^{\mathcal{I}}$, then there is $(x,y) \in r_1^{\mathcal{I}}$ such that $y \in (B \sqcap \bigsqcap_{\exists t.X \in \Psi} \exists t.X)^{\mathcal{I}}$. Since $y \in (B \sqcap \bigsqcap_{\exists t.X \in \Psi} \exists t.X)^{\mathcal{I}}$, there is $(y,z) \in r_2^{\mathcal{I}}$ such that $z \in (C \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X)^{\mathcal{I}}$. We have $(x,y),(y,z) \in s^{\mathcal{I}}$ because $r_1 \sqsubseteq_{\mathcal{T}} s$ and $r_2 \sqsubseteq_{\mathcal{T}} s$. Since $s \circ s \sqsubseteq s \in \mathcal{T}$, $(x,z) \in s^{\mathcal{I}}$. Therefore, $(A \sqcap \bigsqcap_{\exists t.X \in \Phi} \exists t.X) \sqsubseteq_{\mathcal{T}} \exists s.(C \sqcap \bigsqcap_{\exists t.X \in \Omega} \exists t.X)$, i.e., (b) holds.

In the set of rules in Table 5.3, **CR6** and **CR7** add new elements to both E_n and S_n. Thus we need to prove that both (a) and (b) still hold after applying this rule.

CR6 The conditions $\exists s^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$ and $r \sqsubseteq_{\mathcal{T}} s$ lead to $\exists r^-.A_1 \sqsubseteq_{\mathcal{T}} B_1$. Then we can use the proof of soundness for **CF6** in the previous section to prove the soundness property of this rule.

CR7 *(a) holds.*
If $v' \in V_n$, suppose that after applying this rule, $S_{n+1}(v') := S_n(v') \cup \{B_1\}$. We need to prove that $B \sqcap \bigsqcap_{\exists t.X \in \Psi \cup \{\exists r^-.A_1\}} \sqsubseteq_{\mathcal{T}} B_1$. Since $v' := (B, \Psi \cup \{\exists r^-.A_1\})$ and $\exists s^-.A_1 \sqsubseteq B_1 \in \mathcal{T}$, $r \sqsubseteq_{\mathcal{T}} s$, we have $\exists s^-.A_1 \sqsubseteq_{\mathcal{T}} B_1$. Therefore, (a) holds.

If $v' \notin V_n$, then $S_{n+1}(v') := S_n(v) \cup \{B_1\}$. Since $v' := (B, \Psi') = (B, \Psi \cup \{\exists r^-.A_1\})$, we have $(B \sqcap \bigsqcap_{\exists r.X \in \Psi'} \exists r.X) \sqsubseteq_{\mathcal{T}} \exists r^-.A_1$. Thus $(B \sqcap \bigsqcap_{\exists t.X \in \Psi'} \exists t.X) \sqsubseteq_{\mathcal{T}} B_1$. For all $B' \in S_n(v)$, we have $(B \sqcap \bigsqcap_{\exists t.X \in \Psi'} \exists t.X) \sqsubseteq_{\mathcal{T}} (B \sqcap \bigsqcap_{\exists t.X \in \Psi} \exists t.X)$ and $(B \sqcap \bigsqcap_{\exists t.X \in \Psi} \exists t.X) \sqsubseteq B'$. Thus $(B \sqcap \bigsqcap_{\exists t.X \in \Psi'} \exists t.X) \sqsubseteq_{\mathcal{T}} B'$ for all $B' \in S_n(v)$. Therefore, (a) holds.

(b) holds.
Rule **CR7** computes $E_{n+1} = E_n \cup \{(u, r_2, v')\}$.

By induction hypothesis and $(u, r_2, v) \in E_n$, we have $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} \exists r_2.(B \sqcap \bigsqcap_{\exists r.X \in \Psi} \exists r.X)$. Given an $x \in (A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X)^{\mathcal{I}}$ with $\mathcal{I}$ as a model of $\mathcal{T}$, there is a $y \in (B \sqcap \bigsqcap_{\exists r.X \in \Psi} \exists r.X)^{\mathcal{I}}$, such that $(x,y) \in r_2^{\mathcal{I}}$. Together with $r_2 \sqsubseteq_{\mathcal{T}} r$, $(y,x) \in (r^-)^{\mathcal{I}}$. There is $x' \in A_1^{\mathcal{I}}$ such that $(x,x') \in (r_1^-)^{\mathcal{I}}$ because $\exists r_1^-.A_1 \in \Phi$. Together with $r_1 \sqsubseteq_{\mathcal{T}} r$, $(x,x') \in (r^-)^{\mathcal{I}}$.

Since $r \circ r \sqsubseteq r \in \mathcal{T}$, we have $(x', y) \in r^{\mathcal{I}}$. Thus $y \in (B \sqcap \exists r^-.A_1 \sqcap \bigsqcap_{\exists r.X \in \Psi} \exists r.X)^{\mathcal{I}} = (B \sqcap \bigsqcap_{\exists r.X \in \Psi'} \exists r.X)^{\mathcal{I}}$. Therefore, $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} \exists r_2.(B \sqcap \bigsqcap_{\exists r.X \in \Psi'} \exists r.X)$, i.e., (b) holds.

We have finished the proof of Claim 5.3.3.

Using the Claim 5.3.3, it is now easy to prove that $A_0 \sqsubseteq_{\mathcal{T}} B_0$. Let $B_0 \in S(A_0, \emptyset)$. By point (a) of Claim 5.3.3, we have $(A_0 \sqcap \bigsqcap_{\exists r.X \in \emptyset} \exists r.X) \sqsubseteq_{\mathcal{T}} B_0$, i.e., $A_0 \sqsubseteq_{\mathcal{T}} B_0$.

Lemma 5.3.4. *(Completeness) Let (V, E, S) be the completion graph obtained after the application of the rules of Table 5.3 for the normalized general TBox $\mathcal{T}$ has terminated, and let A_0, B_0 be concept names occurring in $\mathcal{T}$. Then $A_0 \sqsubseteq_{\mathcal{T}} B_0$ implies that there exists a node $u = (A_0, \emptyset) \in V$ such that:*

$$B_0 \in S(u)$$

Proof. The lemma is proved by showing the contraposition. Suppose that $B_0 \notin S(u)$ with $u = (A_0, \emptyset)$, we need to prove that this implies $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$. We prove $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ by giving a model $\mathcal{I}$ for the general TBox $\mathcal{T}$, in which there is an $x \in \Delta^{\mathcal{I}}$ such that $x \in A_0^{\mathcal{I}}$ but $x \notin B_0^{\mathcal{I}}$. Before actually defining this model, we define a set E_{bad} of bad edges and prove a claim.

Given a completion graph (V, E, S), the *bad edge set* E_{bad} is defined as the smallest set satisfying the following conditions.

- Case 1: if $(u, r, v) \in E$ and there are $\exists s^-.A \sqsubseteq B \in \mathcal{T}$, $r \sqsubseteq_{\mathcal{T}} s$ with $A \in S(u)$ and $B \notin S(v)$ then $(u, r, v) \in E_{bad}$.
- Case 2: if $(u, r_1, v), (v, r_2, w) \in E$, $v = (B, \Phi)$, $r_1 \sqsubseteq_{\mathcal{T}} r$, $r_2 \sqsubseteq_{\mathcal{T}} r$, $r \circ r \sqsubseteq r \in \mathcal{T}$, $A_1 \in S(u)$ and $\exists r^-.A_1 \notin \Phi$ then $(u, r_1, v) \in E_{bad}$.
- Case 3: if $(u, r_1, v), (v, r_2, w) \in E$, $r_1 \sqsubseteq_{\mathcal{T}} r$, $r_2 \sqsubseteq_{\mathcal{T}} r$, $r \circ r \sqsubseteq r \in \mathcal{T}$, $(u, r_1, v) \in E \backslash E_{bad}$ and $(u, r, w) \in E_{bad}$ then $(v, r_2, w) \in E_{bad}$

This bad edge set is extended from the one for $\mathcal{ELI}$ in [29] by Case 2 and Case 3 in order to preserve the transitive property. We consider the completion graph in Figure 5.4. We have $(u, r, w) \in E_{bad}$ by Case 1, $(u, r_1, v) \in E_{bad}$ by Case 2, and finally $(v', r_2, w) \in E_{bad}$ by Case 3. Therefore, E_{bad} of this graph is $\{(u, r_1, v), (u, r, w), (v', r_2, w)\}$.

Claim 5.3.5. Given the completion graph (V, E, S) and the bad edge set E_{bad}, we have:

If $(u, r, v) \in E_{bad}$ with $v = (A, \Phi)$, then there is
$(u, r, v') \in E \backslash E_{bad}$ with $v' = (A, \Psi)$ and $\Phi \subsetneq \Psi$

The claim is proved for each case:

Case 1: The proof for this case is similar to the one for Claim 5.2.5.

Case 2: If there are $(u, r_1, v), (v, r_2, w) \in E$, $v = (B, \Phi)$, $r_1 \sqsubseteq_{\mathcal{T}} r$, $r_2 \sqsubseteq_{\mathcal{T}} r$, $r \circ r \sqsubseteq r \in \mathcal{T}$, and $(u, r_1, v) \in E_{bad}$ then by **CR6**, a new edge (u, r_1, v'), where $v' = (B, \Psi)$ and $\Phi \subsetneq \Psi$, is added to E.

Case 3: Suppose that $A_1 \in S(u)$ and $v = (B, \Phi)$. Due to the fact that $(u, r_1, v) \in E \backslash E_{bad}$ w.r.t. the condition in Case 1 and Case 2, we have $\exists r^-.A_1 \in \Phi$. Besides, $(v, r_2, w) \in E$, $r_1 \sqsubseteq_{\mathcal{T}} r$, $r_2 \sqsubseteq_{\mathcal{T}} r$ and $r \circ r \sqsubseteq r \in \mathcal{T}$, a new edge (v, r_2, w') that satisfies the claim is added to E by **CR7**.

Therefore, when the algorithm terminates there is a node $v' = (A, \Psi)$ with $\Phi \subsetneq \Psi$ such that (u, r, v') is not a bad edge.

We now define an interpretation $\mathcal{I}$ as follows:

$$\begin{array}{rcl} \Delta^{\mathcal{I}} & := & V; \\ A^{\mathcal{I}} & := & \{u \mid A \in S(u), u \in V\}; \\ s^{\mathcal{I}} & := & \{(u,v) \mid (u,r,v) \in E \backslash E_{bad}, r \sqsubseteq_{\mathcal{T}} s\} \\ & & \cup \{(v,u) \mid (u,r^-,v) \in E \backslash E_{bad}, r \sqsubseteq_{\mathcal{T}} s\}. \end{array}$$

For all $A \in N_{con}$ and $r, s \in N_{role}$.

First we show that there is an $x \in \Delta^{\mathcal{I}}$ such that $x \in A_0^{\mathcal{I}} \backslash B_0^{\mathcal{I}}$, then we need to prove that the interpretation $\mathcal{I}$ is a model of our general TBox $\mathcal{T}$. With the node $u = (A_0, \emptyset)$, $u \in \Delta^{\mathcal{I}}$ we have $B_0 \notin S(u)$ as the supposed hypothesis, i.e., $u \notin B_0^{\mathcal{I}}$. Furthermore, the algorithm starts with $S(A, \emptyset) := \{\top, A\}$ for each $A \in N_{\mathcal{T}}$. So, $A_0 \in S(u)$, i.e., $u \in A_0^{\mathcal{I}}$. Therefore, $(A_0, \emptyset) \in (A_0^{\mathcal{I}} \backslash B_0^{\mathcal{I}})$.

Now we need to prove that $\mathcal{I}$ is a model of $\mathcal{T}$. We make a case distinction according to the form of inclusions axioms.

- $A \sqsubseteq B$.
  Let $u \in A^{\mathcal{I}}$, by the definition of $A^{\mathcal{I}}$, we have $A \in S(u)$. Due to **CR1**, this implies $B \in S(u)$, thus $u \in B^{\mathcal{I}}$.

- $A_1 \sqcap A_2 \sqsubseteq B$.
  Let $u \in A_1^{\mathcal{I}} \sqcap A_2^{\mathcal{I}}$, by the definition of $A_1^{\mathcal{I}}$, $A_2^{\mathcal{I}}$, we have $A_1, A_2 \in S(u)$. Due to **CR2**, this implies $B \in S(u)$, thus $u \in B^{\mathcal{I}}$.

- $A \sqsubseteq \exists r.B$.
  Let $u \in A^{\mathcal{I}}$, by the definition of $A^{\mathcal{I}}$, we have $A \in S(u)$. Due to **CR3**, there exists $(u, r, v) \in E$ with $v = (B, \emptyset)$. There are two cases:

  - *First case:* $(u, r, v) \in E \backslash E_{bad}$: From the definition of $r^{\mathcal{I}}$, we have $(u, v) \in r^{\mathcal{I}}$. Because $B \in S(B, \emptyset)$, we have $v \in B^{\mathcal{I}}$. Together, this yields $u \in (\exists r.B)^{\mathcal{I}}$.
  - *Second case:* $(u, r, v) \in E_{bad}$: From Claim 5.3.5, there is $v' = (B, \Phi) \neq v$ such that $(u, r, v') \in E \backslash E_{bad}$. Therefore $(u, v') \in r^{\mathcal{I}}$. Since $B \in S(B, \Phi)$ for all Φ, we have $v' \in B^{\mathcal{I}}$. Together, this yields $u \in (\exists r.B)^{\mathcal{I}}$.

- $\exists s.A \sqsubseteq B$.
  Let $u \in (\exists s.A)^{\mathcal{I}}$, which means that there is $v \in A^{\mathcal{I}}$ such that $(u, v) \in r^{\mathcal{I}}$. According to the definition of $r^{\mathcal{I}}$ above, there is $r \in N_{role}$ such

that $r \sqsubseteq_{\mathcal{T}} s$ and $(u, r, v) \in E$ or $(v, r^-, u) \in E$. We consider these two possibilities:

- *First case:* $(u, r, v) \in E \backslash E_{bad}$: Due to Rule **CR4**, we have $B \in S(u)$, which means that $u \in B^{\mathcal{I}}$.
- *Second case:* $(v, r^-, u) \in E \backslash E_{bad}$: As we know the relation between r and r^-:

$$(u, v) \in r^{\mathcal{I}} \text{ iff } (v, u) \in (r^-)^{\mathcal{I}}$$

We rename $r' = r^-$, i.e., $r = (r')^-$ and $s' = s^-$, i.e., $s = (s')^-$ to make its look more appropriate:

$(v, r', u) \in E \backslash E_{bad}$, $\exists (s')^-.A \sqsubseteq B \in \mathcal{T}$, $A \in S(v)$, $r' \sqsubseteq_{\mathcal{T}} s'$.

Since $(v, r', u) \notin E_{bad}$, we have $B \in S(u)$, which means that $u \in B^{\mathcal{I}}$.

- $r \sqsubseteq s$

  This axiom holds because of the definition of roles' interpretation.

- $r \circ r \sqsubseteq r$.

  Due to **CR5**, the transitive property holds in the completion graph (V, E, S). This property is not affected by removing bad edges from the graph due to the definition of E_{bad}. Together with the definition of the interpretation, this axiom holds.

Therefore the completeness of the algorithm is proved.

5.4 An algorithm for $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes

In this section function f is redefined as follows: $N_{role} \rightarrow \{True, False\}$ such that $f(r)$ iff $\top \sqsubseteq (\leq 1s) \in \mathcal{T}$ and $r \sqsubseteq_{\mathcal{T}} s$.

The set of completion rules in the algorithm for $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes includes all the rules in Table 5.3 except **CR3** and additional rules in Table 5.4. **CFR1** and **CFR2** are similar to the ones for $\mathcal{ELI}f$. Due to **CFR1**, if there is $A \sqsubseteq \exists r.B \in \mathcal{T}$ with $f(r)$, the node $(\top, \{\exists r^-.A\})$ is initialized with $S(\top, \{\exists r^-.A\}) = B$.

The purpose of **CFR3** is to deal with the interaction between role hierarchy and functional roles.

In Table 5.4, A, B, A_1, B_1 are concept names or top; u, v are nodes in the graph; Ψ and $\Omega$are sets of concept descriptions of the form $\exists r.A$, and r is either a role or inverse role. In our algorithm, rules in Table 5.3 except **CR3** and in Table 5.4 are applied if and only if their application changes the completion graph.

CFR1	If $A \in S(u)$, $v = (B, \emptyset)$,$w = (\top, \{\exists r^-.A\})$, $A \sqsubseteq \exists r.B \in \mathcal{T}$ then if $f(r)$ then $E := E \cup \{(u, r, w)\}$; $S(w) := S(w) \cup \{B_1\}$ else $E := E \cup \{(u, r, v)\}$
CFR2	If $(u, r_1, v) \in E$, $B \in S(v)$, $r_1 \sqsubseteq_{\mathcal{T}} s$, $r_2 \sqsubseteq_{\mathcal{T}} s$ and $B \sqsubseteq \exists r_2^-.A \in \mathcal{T}$, $f(s^-)$ then $S(u) := S(u) \cup \{A\}$
CFR3	If $(u, r_1, v), (u, r_2, w) \in E$, $r_1 \sqsubseteq_{\mathcal{T}} s$, $r_2 \sqsubseteq_{\mathcal{T}} s$, $f(s)$ and $v = (\top, \Psi)$, $w = (\top, \Omega)$, $A_1 \in S(u)$ then $v' := (\top, \Psi \cup \Omega)$; if $v' \notin V$ then $V := V \cup \{v'\}$ and $S(v') := S(v) \cup S(w)$, else $S(v') := S(v') \cup S(v) \cup S(w)$; $E := E \cup \{(u, r_1, v')\}$

Table 5.4: Additional Completion Rules for $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes

$A \sqsubseteq \exists s_1.B$	$s_1 \sqsubseteq s$	$\top \sqsubseteq (\leq 1 s^-)$
$B \sqsubseteq \exists s_2^-.X$	$s_2 \sqsubseteq s$	$\top \sqsubseteq (\leq 1 r)$
$X \sqsubseteq \exists r_1.B_1$	$r_1 \sqsubseteq r$	$t \circ t \sqsubseteq t$
$X \sqsubseteq \exists r_2.B_2$	$r_2 \sqsubseteq r$	
$B_1 \sqcap B_2 \sqsubseteq C$	$s_1 \sqsubseteq t$	
$\exists r.C \sqsubseteq Y$	$s_2^- \sqsubseteq t$	

Figure 5.5: An $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox

We consider an example $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox as in Figure 5.5.

Applying the set of completion rules in Table 5.4 we get a completion graph in Figure 5.6. We consider a notable **CFR3** application. After **CFR1** applications we have (u, r_1, x) and (u, r_2, y), which is then applied to **CFR3** and get (u, r_1, x').

In the next part we prove that the algorithm terminates, is sound and complete.

We start with a lemma about the algorithm's termination, which is proved similarly to the one with $\mathcal{ELHI}_{\mathcal{R}^+}$ in the previous section.

Lemma 5.4.1. *For a normalized $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox $\mathcal{T}$, the algorithm runs in exponential time.*

Lemma 5.4.2. *(Soundness) Let (V, E, S) be the completion graph obtained after the application of the rules from Table 5.3 except **CR3** and in Table 5.4*

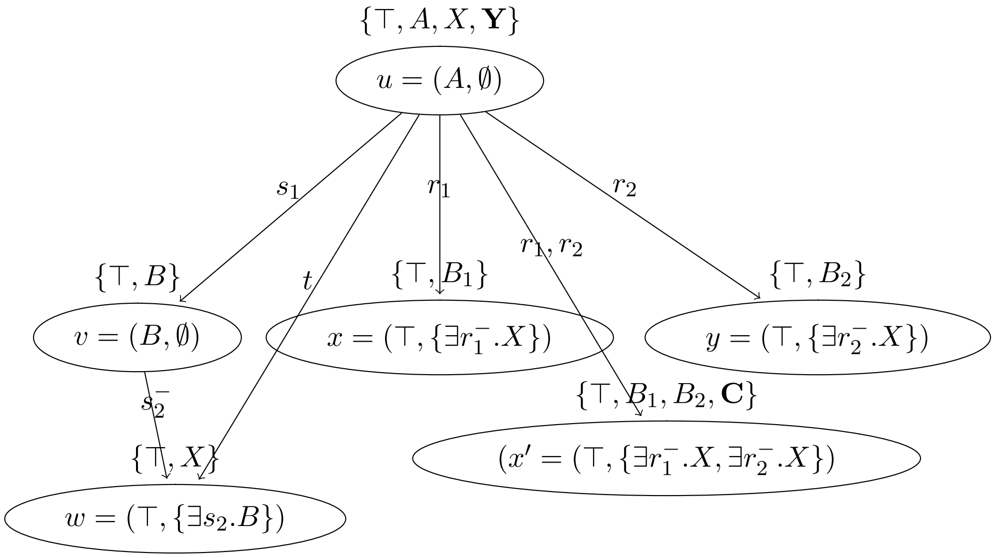

Figure 5.6: An example of reasoning in $\mathcal{ELHI}f_{\mathcal{R}^+}$

on the normalized $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBox $\mathcal{T}$ has terminated, and let A_0, B_0 be two concept names occurring in $\mathcal{T}$. Then $A_0 \sqsubseteq_{\mathcal{T}} B_0$ if the following condition holds:

$$B_0 \in S(u) \text{ with } u = (A_0, \emptyset) \in V$$

Proof. Let $(V_0, E_0, S_0), \ldots, (V_{n_0}, E_{n_0}, S_{n_0})$ be the sequence of description graphs produced by the algorithm. Assume that $B_0 \in S(u)$ with $u = (A_0, \emptyset) \in V$. Before proving that $A_0 \sqsubseteq_{\mathcal{T}} B_0$, we prove the following claim.

Claim 5.4.3. Given the general TBox $\mathcal{T}$, assume that there are two nodes $u = (A, \Phi)$, $v = (B, \Psi)$, an edge r in the completion graph (V, E, S) and a concept name C.

(a) If $C \in S(A, \Phi)$, then $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} C$; and

(b) if $((A, \Phi), r, (B, \Psi)) \in E$,
then $(A \sqcap \bigsqcap_{\exists s.X \in \Phi} \exists s.X) \sqsubseteq_{\mathcal{T}} \exists r.(B \sqcap \bigsqcap_{\exists s.X \in \Psi} \exists s.X)$

The claim is proved by induction on $n \in \{0, \ldots, n_0 - 1\}$.

For the induction start, $n = 0$ implies $S_n(v) := \{\top, A\}$ for all $v = (A, \emptyset)$, $V_n := \{(A, \emptyset) | A \in N_{\mathcal{T}}\}$, $E_n := \emptyset$. For (a), $C \in S_0(A, \Phi)$ implies $C = \top$ or $C = A$, thus $(A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X) \sqsubseteq_{\mathcal{T}} C$. Since $E_0 = \emptyset$, $((A, \Phi), r, (B, \Psi)) \in E$ is always false, thus (b) holds.

For the induction step, we make a case distinction according to the rule used to add a new concept to $S_n(v)$ for $v \in V_n$ or a new edge to E_n. The proofs for rules in Table 5.3 are given in the previous section. The proofs for **CFR1** and **CFR2** are similar to **CF3** and **CF5** in Table 5.2

The last completion rule **CFR3** is given to support the combination of role hierarchies and functional roles.

We only need to prove that both the conditions hold for **CFR3**.

(a) holds.
Given $B_1 \in S_n(v) \cup S_n(w)$, we need to prove that $(\top \sqcap \bigsqcap_{\exists r.X \in \Psi} \exists r.X) \sqcap (\top \sqcap \bigsqcap_{\exists r.X \in \Omega} \exists r.X) \sqsubseteq_{\mathcal{T}} B_1$.
If $B_1 \in S_n(u)$, by (a), we have $\top \sqcap \bigsqcap_{\exists r.X \in \Psi} \exists r.X \sqsubseteq B_1$.
If $B_1 \in S_n(w)$, by (a), we have $\top \sqcap \bigsqcap_{\exists r.X \in \Omega} \exists r.X \sqsubseteq B_1$.
Therefore, (a) holds.

(b) holds.
Suppose that $x \in (A \sqcap \bigsqcap_{\exists r.X \in \Phi} \exists r.X)^{\mathcal{I}}$ with $u = (A, \Phi)$, then there exist $y \in (\top \sqcap \bigsqcap_{\exists r.X \in \Psi} \exists r.X)^{\mathcal{I}}$ and $z \in (\top \sqcap \bigsqcap_{\exists r.X \in \Omega} \exists r.X)^{\mathcal{I}}$ such that $(x, y) \in r_1^{\mathcal{I}}$, $(x, z) \in r_2^{\mathcal{I}}$. Since $r_1 \sqsubseteq_{\mathcal{T}} s$ and $r_2 \sqsubseteq_{\mathcal{T}} s$ we have $(x, y), (x, z) \in s^{\mathcal{I}}$. Together with $f(s)$, we have $y = z$. Therefore, $y \in (\top \sqcap \bigsqcap_{\exists r.X \in \Psi} \exists r.X)^{\mathcal{I}} \cap (\top \sqcap \bigsqcap_{\exists r.X \in \Omega} \exists r.X)^{\mathcal{I}}$, i.e., $y \in (\top \sqcap \bigsqcap_{\exists r.X \in \Psi} \exists r.X \sqcap \bigsqcap_{\exists r.X \in \Omega} \exists r.X)^{\mathcal{I}}$.
Thus, (b) holds.

We have finished the proof of Claim 5.4.3.

Using the Claim 5.4.3, it is now easy to prove that $A_0 \sqsubseteq_{\mathcal{T}} B_0$. Let $B_0 \in S(A_0, \emptyset)$. By point (a) of Claim 5.4.3, we have $(A_0 \sqcap \bigsqcap_{\exists r.X \in \emptyset} \exists r.X) \sqsubseteq_{\mathcal{T}} B_0$, i.e., $A_0 \sqsubseteq_{\mathcal{T}} B_0$.

Lemma 5.4.4. *(Completeness) Let (V, E, S) be the completion graph obtained after the application of the rules from Table 5.4 on the normalized general TBox $\mathcal{T}$ has terminated, and let A_0, B_0 be concept names occurring in $\mathcal{T}$. Then $A_0 \sqsubseteq_{\mathcal{T}} B_0$ implies that there exists a node $u = (A_0, \emptyset) \in V$ such that:*

$$B_0 \in S(u)$$

Proof. The lemma is proved by showing the contraposition. Suppose that $B_0 \notin S(u)$ with $u = (A_0, \emptyset)$, we need to prove that this implies $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$. We prove $A_0 \not\sqsubseteq_{\mathcal{T}} B_0$ by giving a counter model $\mathcal{I}$ for the general TBox $\mathcal{T}$, in which there is $x \in \Delta^{\mathcal{I}}$ such that $x \in A_0^{\mathcal{I}}$ but $x \notin B_0^{\mathcal{I}}$.

This counter model is based on a sub-graph of our graph (V, E, S), which is constructed below.

Firstly, we define a set E_{bad} of bad edges as in the completeness proof of $\mathcal{ELHI}_{\mathcal{R}^+}$ and prove a claim.

Given a completion graph (V, E, S), the *bad edge set* E_{bad} is defined as the smallest set satisfying the following.

- Case 1: If $(u, r, v) \in E$ and there are $\exists s^-.A \sqsubseteq B, r \sqsubseteq_\mathcal{T} s$ such that $A \in S(u)$ and $B \notin S(v)$ then $(u, r, v) \in E_{bad}$.
- Case 2: If $(u, r_1, v), (v, r_2, w) \in E$, $v = (B, \Psi)$, $r_1 \sqsubseteq_\mathcal{T} r$, $r_2 \sqsubseteq_\mathcal{T} r$, $r \circ r \sqsubseteq r \in \mathcal{T}$, $A_1 \in S(u)$ and $\exists r^-.A_1 \notin \Psi$ then $(u, r_1, v) \in E_{bad}$.
- Case 3: If $(u, r_1, v), (v, r_2, w) \in E$, $r_1 \sqsubseteq_\mathcal{T} r$, $r_2 \sqsubseteq_\mathcal{T} r$, $r \circ r \sqsubseteq r \in \mathcal{T}$, $(u, r_1, v) \in E \backslash E_{bad}$ and $(u, r, w) \in E_{bad}$ then $(v, r_2, w) \in E_{bad}$

Claim 5.4.5. Given the completion graph (V, E, S) and the bad edge set E_{bad}, we have:

If $(u, r, v) \in E_{bad}$ with $v = (A, \Phi)$, then there is
$(u, r, v') \in E \backslash E_{bad}$ with $v' = (A, \Psi)$ and $\Phi \subsetneq \Psi$

The claim is proved similar to the same one in the section of $\mathcal{ELHI}_{\mathcal{R}^+}$.

For a node u in the graph, we define an equivalence relation $\sim_{u,\mathcal{T}}$.

$r_1 \sim_{u,\mathcal{T}} r_2$ iff there is an $r \in N_{role}$ such that $r_1 \sqsubseteq_\mathcal{T} r$, $r_2 \sqsubseteq_\mathcal{T} r$ and $f(r)$ and there exist two edge (u, r_1, v_1) and (u, r_2, v_2) in the graph.

The set $\tilde{r^u}$ is inductively defined as follows.

$\tilde{r^u}_0 := \{r\}$ if there exists $(u, r, v) \in E$ for some $v \in V$;

$\tilde{r^u}_{i+1} := \tilde{r^u}_i \cup \{s' \mid s \in \tilde{r^u}_i, s' \sim_{u,\mathcal{T}} s\}$

until $\tilde{r^u}_{n-1} = \tilde{r^u}_n = \tilde{r^u}$.

Assuming that $u \in V$, $\top \sqsubseteq (\leq 1r) \in \mathcal{T}$, then for all $(u, s, v) \in E$ such that $s \in \tilde{r^u}$, we define the same node $\tau(u, r)$, having the following property:

Claim 5.4.6. $(u, s, \tau(u, r)) \in E \backslash E_{bad}$ and if $A \in S(u)$, $A \sqsubseteq \exists s.B \in \mathcal{T}$ then $B \in S(\tau(u, r))$.

The Claim is proved by considering the applications of **CFR1, CFR2, CFR3**.

We inductively define the sub-graph (V^*, E^*, S) of (V, E, S) as follows:

$V_0^* = \{(A_0, \emptyset)\}$;

If there is $(u, r, v) \in E \backslash E_{bad}$ such that $u \in V_i^*$ then if $\neg f(r)$ then $V_{i+1}^* := V_i^* \cup \{v\}$; $E_{i+1}^* := E_i^* \cup \{(u, r, v)\}$

else if $f(r)$ and there is $(v', r^-, u) \in E_i^*$ then $V_{i+1}^* := V_i^*$; $E_{i+1}^* := E_i^* \cup \{(u, r, v')\}$

else $V_{i+1}^* := V_i^* \cup \{\tau(u, r)\}$; $E_{i+1}^* := E_i^* \cup \{(u, r, \tau(u, r))\}$

until the defining graph cannot change.

For instance, given a completion graph (V, E, S) as in Figure 5.6, we have:

$$V^* = \{u, v, w, x'\};$$

$$E^* = \{(u, s_1, v), (v, s_2, u), (u, t, w), (u, r_1, x'), (u, r_2, x')\}$$

We now define a model $\mathcal{I}$ based on (V^*, E^*, S) as follows:

$$\begin{aligned}
\Delta^{\mathcal{I}} &:= V^*; \\
A^{\mathcal{I}} &:= \{u \mid A \in S(u), u \in V^*\}; \\
s^{\mathcal{I}} &:= \{(u, v) \mid (u, r, v) \in E^*, r \sqsubseteq_{\mathcal{T}} s\} \cup \{(v, u) \mid (u, r^-, v) \in E^*, r \sqsubseteq_{\mathcal{T}} s\}.
\end{aligned}$$

For all $A \in N_{con}$ and $r, s \in N_{role}$.

Then we unravel this model with the root at $(A_0, \emptyset)$ and get a possibly infinite tree model $\mathcal{I}$.

Firstly, we show that there is $x \in \Delta^{\mathcal{I}}$ such that $x \in A_0^{\mathcal{I}} \backslash B_0^{\mathcal{I}}$, then we need to prove that the interpretation $\mathcal{I}$ is a model of our general TBox $\mathcal{T}$. With the node $u = (A_0, \emptyset)$, $u \in \Delta^{\mathcal{I}}$ we have $B_0 \notin S(u)$ as the supposed hypothesis, i.e., $u \notin B_0^{\mathcal{I}}$. Furthermore, the algorithm starts with $S(A, \emptyset) := \{\top, A\}$ for each $A \in N_{\mathcal{T}}$. So, $A_0 \in S(u)$, i.e., $u \in A_0^{\mathcal{I}}$. Therefore, $(A_0, \emptyset) \in (A_0^{\mathcal{I}} \backslash B_0^{\mathcal{I}})$.

Now we need to prove that $\mathcal{I}$ is a model of $\mathcal{T}$. We make a case distinction according to the form of concept inclusions.

- $A \sqsubseteq B$.
  Let $u \in A^{\mathcal{I}}$, by the definition of $A^{\mathcal{I}}$, we have $A \in S(u)$. Due to Rule **CR1**, this implies $B \in S(u)$, thus $u \in B^{\mathcal{I}}$.

- $A_1 \sqcap A_2 \sqsubseteq B$.
  Let $u \in A_1^{\mathcal{I}} \sqcap A_2^{\mathcal{I}}$, by the definition of $A_1^{\mathcal{I}}$, $A_2^{\mathcal{I}}$, we have $A_1, A_2 \in S(u)$. Due to Rule **CR2**, this implies $B \in S(u)$, thus $u \in B^{\mathcal{I}}$.

- $A \sqsubseteq \exists r.B$.
  Let $u \in A^{\mathcal{I}}$, by the definition of $A^{\mathcal{I}}$, we have $A \in S(u)$.

Due to the definition of $\mathcal{I}$ and **CFR1**, there exists $(u, r, v) \in E$.

If r is not a functional role, then there are two cases:

- *First case:* $(u, r, v) \in E \backslash E_{bad}$: From the definition of $r^{\mathcal{I}}$, we have $(u, v) \in r^{\mathcal{I}}$. Since $B \in S(B, \emptyset)$, we have $v \in B^{\mathcal{I}}$. Together, this yields $u \in (\exists r.B)^{\mathcal{I}}$.
- *Second case:* $(u, r, v) \in E_{bad}$: From Claim 5.4.5, there is $v' = (B, \Phi) \neq v$ such that $(u, r, v') \in E \backslash E_{bad}$. Therefore $(u, v') \in r^{\mathcal{I}}$. Since $B \in S(B, \Phi)$ for all Φ, we have $v' \in B^{\mathcal{I}}$. Together, this yields $u \in (\exists r.B)^{\mathcal{I}}$.

If r is a functional role, we need to check two cases:

- If there is $u' \in \Delta^{\mathcal{I}}$ such that $(u', u) \in (r^-)^{\mathcal{I}}$: By **CR6** application, we have $B \in S(u')$, i.e., $u' \in B^{\mathcal{I}}$. Since $(u, u') \in r^{\mathcal{I}}$, this axiom holds.
- If there is not any $u' \in \Delta^{\mathcal{I}}$ such that $(u', u) \in (r^-)^{\mathcal{I}}$: by Claim 5.4.6, this axiom holds.

- $\exists s.A \sqsubseteq B$.
Let $u \in (\exists s.A)^{\mathcal{I}}$, which means that there is $v \in A^{\mathcal{I}}$ such that $(u, v) \in r^{\mathcal{I}}$. According to the definition of $r^{\mathcal{I}}$ above, there is $r \in N_{role}$ such that $r \sqsubseteq_{\mathcal{T}} s$ and $(u, r, v) \in E^*$ or $(v, r^-, u) \in E^*$. We consider these two possibilities:
  - *First case:* $(u, r, v) \in E^*$: Due to Rule **CR4**, we have $B \in S(u)$, which means that $u \in B^{\mathcal{I}}$.
  - *Second case:* $(v, r^-, u) \in E^*$: As we know the relation between r and r^-:

$$(u, v) \in r^{\mathcal{I}} \text{ iff } (v, u) \in (r^-)^{\mathcal{I}}$$

    We rename $r' = r^-$, i.e., $r = (r')^-$ and $s' = s^-$, i.e., $s = (s')^-$ to make its look more appropriate:
    $(v, r', u) \in E^*$, $\exists (s')^-.A \sqsubseteq B \in \mathcal{T}$, $A \in S(v)$, $r' \sqsubseteq_{\mathcal{T}} s'$.
    Since $(v, r', u) \notin E_{bad}$, we have $B \in S(u)$, which means that $u \in B^{\mathcal{I}}$.

- $\top \sqsubseteq (\leq 1 r)$.
This axiom holds by the way we construct the model.

- $r \sqsubseteq s$
This axiom holds because of the definition of roles' interpretation.

- $r \circ r \sqsubseteq r$.
  Due to **CR5**, the transitive property holds in the completion graph (V, E, S). This property is not affected by removing bad edges from the graph due to the definition of E_{bad}. Together with the definition of the interpretation, this axiom holds. Since transitive roles are not functional ones, transitive property is not affected when we modify the graph in order to satisfy functional axioms.

  Therefore, this axiom holds.

Thus the completeness of the algorithm is proved.

5.5 A special case with $\mathcal{ELH}_{\mathcal{R}^+}$

CR1'	If $A \in S(v)$, $A \sqsubseteq B \in \mathcal{T}$ then $S(v) := S(v) \cup \{B\}$
CR2'	If $A_1, A_2 \in S(v)$, $A_1 \sqcap A_2 \sqsubseteq B \in \mathcal{T}$ then $S(v) := S(v) \cup \{B\}$
CR3'	If $A \in S(u)$, $v = (B, \emptyset)$, $A \sqsubseteq \exists r.B \in \mathcal{T}$ then $E := E \cup \{(u, r, v)\}$
CR4'	If $(u, r, v) \in E$, $B \in S(v)$, $\exists s.B \sqsubseteq A \in \mathcal{T}$, $r \sqsubseteq_{\mathcal{T}} s$ then $S(u) := S(u) \cup \{A\}$
CR5'	If $(u, r_1, v), (v, r_2, w) \in E$, $r_1 \sqsubseteq_{\mathcal{T}} s$, $r_2 \sqsubseteq_{\mathcal{T}} s$, $s \circ s \sqsubseteq s \in \mathcal{T}$ then $E := E \cup \{(u, s, w)\}$

Table 5.5: Completion Rules for $\mathcal{ELH}_{\mathcal{R}^+}$-TBoxes

In the case the input $\mathcal{ELHI}_{\mathcal{R}^+}$ general TBoxes do not have inverse roles, they become $\mathcal{ELH}_{\mathcal{R}^+}$ general TBoxes and the algorithm of $\mathcal{ELHI}_{\mathcal{R}^+}$ becomes polynomial.

There is no inverse role in the description logic $\mathcal{ELH}_{\mathcal{R}^+}$, therefore conditions related to inverse roles in "branching" rules **CR6** and **CR7** can not be satisfied. Thus the algorithm does not use these two rules when it works in the DL $\mathcal{ELH}_{\mathcal{R}^+}$.

All the nodes start with the form $(A, \emptyset)$, where A is a concept name. When the rules **CR6** and **CR7** are not applied, these nodes keep that form until the algorithm terminates. We rename nodes having the form $u = (A, \emptyset)$

to A, then the set of rules is shown in Table 5.5, which is similar to the set of rules for the DL $\mathcal{EL}^+$ in [9] . Therefore the algorithm works polynomially for the DL $\mathcal{ELH}_{\mathcal{R}^+}$. We end the chapter by the last theorem, which is an interesting point of the proposed algorithm.

Theorem 5.5.1. *The algorithm works in polynomial time for the case of the DL $\mathcal{ELH}_{\mathcal{R}^+}$ w.r.t. general TBoxes.*

Chapter 6

Conclusion

In this thesis, we have proposed methods for deciding the subsumption in the DL $\mathcal{ELHI}f_{\mathcal{R}^+}$, which is sufficient to formulate the GALEN ontology. Some tractable extensions of DL $\mathcal{EL}$ have been used for this ontology. However, these tractable $\mathcal{EL}$-extensions can only represent parts of GALEN ontology. The thesis has proposed non-tableau-based solutions for the subsumption problem in $\mathcal{ELHI}f_{\mathcal{R}^+}$ general TBoxes.

The solutions for this problem lie in two directions. The first direction is attempting to reduce the input TBoxes to ones formulated in a less expressive DL. The reduction stops when $\mathcal{ELI}$ general TBoxes, which can be classified using the previous algorithm in [29], are obtained. The other direction to solve subsumption in these DLs is building sets of completion rules for the algorithms. Even though these two directions share some basic ideas, the latter direction offers direct advantages both in theory and in practice.

This algorithm runs in polynomial time if the input TBox is formulated in the sub-language of the DL $\mathcal{ELH}_{\mathcal{R}^+}$. Since the inverse roles are the main reason to make the algorithm complicated, when the number of inverse roles in the ontology is small, the algorithm's complexity can be significantly reduced. Thus it is expected that the algorithm will perform better than tableaux-based algorithms in practice.

Another advantage of the algorithm is that it classifies the ontology. In other tableau-based DL systems, the subsumption hierarchy is computed through multiple subsumption tests between pairs of concept names. However, in this algorithm, subsumption between all pairs of concepts in the input TBox is checked simultaneously.

One of the future directions that should be considered is optimizing and implementing these algorithms to see their performance in practice. The sets of completion rules at present do not appear simple but they are useful for implementation.

Another direction is to make it work well with other intractable extensions of $\mathcal{EL}$, though the value of them in practice should also be considered.

There are some extensions that are easily reduced to $\mathcal{EL}$-extensions with inverse roles, such as the $\mathcal{EL}$-extension having symmetric roles. However, there are extensions that need more effort to make the current algorithms work with them.

It is believed that the proposed algorithm can be extended to solve other problems in $\mathcal{EL}$-extensions, such as *Pinpointing in the Description Logic.* A pinpointing algorithm that is similar to the one shown in [10] for $\mathcal{EL}$, except that our algorithm works with dynamic completion graphs.

Bibliography

[1] F. Baader. Using automata theory for characterizing the semantics of terminological cycles. *Annals of Mathematics and Artificial Intelligence*, 18(2–4):175–219, 1996.

[2] F. Baader, S. Brandt, and C. Lutz. Pushing the $\mathcal{EL}$ envelope. LTCS-Report LTCS-05-01, Chair for Automata Theory, Institute for Theoretical Computer Science, Dresden University of Technology, Germany, 2005. 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, 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.

[5] F. Baader and W. Nutt. Basic description logics. In Franz Baader, Diego Calvanese, Deborah McGuinness, Daniele Nardi, and Peter F. Patel-Schneider, editors, *The Description Logic Handbook: Theory, Implementation, and Applications*, pages 43–95. Cambridge University Press, 2003.

[6] 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.

[7] Franz Baader, Sebastian Brandt, and Carsten Lutz. Pushing the $\mathcal{EL}$ envelope further. In Kendall Clark and Peter F. Patel-Schneider, editors, *In Proceedings of the OWLED 2008 DC Workshop on OWL: Experiences and Directions*, 2008.

[8] 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.

[9] 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.

[10] Franz Baader, Rafael Peñaloza, and Boontawee Suntisrivaraporn. Pinpointing in the description logic $\mathcal{EL}$. In *Proceedings of the 2007 International Workshop on Description Logics (DL2007)*, CEUR-WS, 2007.

[11] Ronald J. Brachman and Hector J. Levesque. The tractability of subsumption in frame-based description languages. In *AAAI*, pages 34–37, 1984.

[12] 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.

[13] The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat Genet*, 25(1):25–9, 2000.

[14] Volker Haarslev and Ralf Möller. Racer system description. In *IJCAR '01: Proceedings of the First International Joint Conference on Automated Reasoning*, pages 701–706, London, UK, 2001. Springer-Verlag.

[15] Ian Horrocks. *Optimising Tableaux Decision Procedures for Description Logics*. PhD thesis, University of Manchester, 1997.

[16] Ian Horrocks. Using an expressive description logic: FaCT or fiction? In *Proc. of the 6th Int. Conf. on Principles of Knowledge Representation and Reasoning (KR'98)*, pages 636–647, 1998.

[17] Ian Horrocks and Peter Patel-Schneider. Reducing OWL entailment to description logic satisfiability. *J. of Web Semantics*, 1(4):345–357, 2004.

[18] Ian Horrocks, Ulrike Sattler, and Stephan Tobies. Practical reasoning for expressive description logics. In *LPAR '99: Proceedings of the 6th International Conference on Logic Programming and Automated Reasoning*, pages 161–180, London, UK, 1999. Springer-Verlag.

[19] Sherrilynne S. Fuller Hsinchun Chen. *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*. Springer, 2005.

[20] Yevgeny Kazakov and Hans de Nivelle. Subsumption of concepts in $\mathcal{FL}_0$ with respect to descriptive semantics is pspace-complete.

[21] Marvin Minsky. A framework for representing knowledge. Technical report, Cambridge, MA, USA, 1974.

[22] Bernhard Nebel. Computational complexity of terminological reasoning in back. *Artif. Intell.*, 34(3):371–383, 1988.

[23] M. Rose Quillian. Word concepts: A theory and simulation of some basic semantic capabilities. *Behavioral Science*, 1967.

[24] A. Rector, S. Bechhofer, C. A. Goble, I. Horrocks, W. A. Nowlan, and W. D. Solomon. The GRAIL concept modelling language for medical terminology. *Artificial Intelligence in Medicine*, 9:139–171, 1997.

[25] A. L. Rector, A. Gangemi, E. Galeazzi, A. J. Glowinski, and A. Rossi-Mori. The galen core model schemata for anatomy: Towards a re-usable application-independent model of medical concepts. *Twelfth International Congress of the European Federation for Medical Informatics, MIE-94, Lisbon, Portugal*, 1994.

[26] A. L. Rector and J. E. Rogers. Ontological issues in using a description logic to represent medical concepts: Experience from galen. *Methods of Information in Medicine*, (In Press), 2000.

[27] A. L. Rector and J. E. Rogers. Ontological and practical issues in using a description logic to represent medical concept systems: Experience from galen. 4126/2006, 2006.

[28] K. A. Spackman. Managing clinical terminology hierarchies using algorithmic calculation of subsumption: Experience with SNOMED-RT. In *J. of the American*, 2000. Fall Symposium Special Issue.

[29] Huy Q. Vu. Reasoning in $\mathcal{ELI}$ with GCIs. Project report, TU Dresden, Germany, 2008.